Tytuł: **Koncepcja umysłu w filozofii Daniela C. Dennetta**

Autor: Piotr Czarnecki; doandpet@poczta.onet.pl

Skład i rysunki: Marek Kasperski, wg oryginału

Recenzent: prof. dr hab. Urszula Żegleń (UMK) i prof. dr hab. Włodzisław Duch (UMK)

Źródło: http://kognitywistyka.prv.pl/; mcmarcus@poczta.onet.pl

0. Wstęp

Filozofia umysłu poszukuje odpowiedzi na pytania dotyczące fenomenu świadomości, możliwości istnienia świadomych maszyn, świadomości u zwierząt, myślenia, percepcji, zależności pomiędzy umysłem a tradycyjnie pojmowaną materią oraz wielu innych. Klasycznym problemem jest problem wzajemnej relacji umysłu i ciała [*mind-body problem*]. Zastąpiony on został, postawionym bardziej ogólnie, problemem relacji między tym, co mentalne a tym, co fizyczne. Dotyczy on teraz relacji pomiędzy umysłem i mózgiem [*mind-brain problem*] oraz innych podobnych relacji. Problem ten będzie punktem wyjścia dla naturalistycznej koncepcji Daniela Dennetta. Celem tej pracy jest przedstawienie głoszonej przez niego koncepcji umysłu. Wykazane ma zostać również, że koncepcja ta, wraz ze wszystkimi jej nawiązaniami do nauk pokrewnych, stanowi dobry, choć tylko ogólny, przegląd współczesnego stanu wiedzy na temat mózgu, umysłu, cybernetyki i sztucznej inteligencji. Za aktualnością pracy przemawia to, że dotyczy ona wielu podejmowanych współcześnie problemów. Dennett nie podaje dla nich ostatecznych rozwiązań, ani nie wnika w szczegóły neurofizjologii czy genetyki. Nakreśla tylko ogólnie kierunek przyszłych badań, a praca ta ma kierunek ten naszkicować i spróbować go ocenić.

Przyjęta metoda polega na wyodrębnieniu czterech grup problemów, z których każda omawiana jest w jednym z rozdziałów. Podział taki umożliwi pokazanie jak Dennett wychodząc od filozofii, stara się połączyć w swej holistycznej koncepcji, naturę i organizmy żywe, z jednej strony, z kulturą i takimi jej wytworami jak maszyny cyfrowe, z drugiej. Podzielenie podejmowanych przez Dennetta problemów na cztery grupy tematyczne ma służyć pokazaniu, z jakich założeń filozoficznych on korzysta, jakie wyznaczył miejsce dla intencjonalności w procesie ewolucji, jak przedstawia świadomy umysł, jako ostatnie stadium ewolucyjne systemów intencjonalnych oraz jak pełną niejasności, lecz mimo tego obiecującą dziedziną jest teoria sztucznej inteligencji.

Rozdział pierwszy dotyczy podstaw filozoficznych i założeń przyjmowanych w koncepcji Dennetta. Główne znaczenie ma tutaj intencjonalność, która potraktowana zostaje jako cecha

istotna umysłów. Dla rozpoznania intencjonalności wykorzystana zostaje referencyjność oraz kryteria Chisholma służące rozpoznaniu zdań intencjonalnych. Intencjonalność omawiana jest z perspektywy ewolucji i organizmów żywych, które zostają nazwane systemami intencjonalnymi. Skrytykowany zostaje behawioryzm, gdyż intencjonalność omawianych systemów może być stwierdzana tylko wewnętrznie. Zreferowane zostaje wprowadzone przez Dennetta nastawienie fizykalne, projektowe i intencjonalne.

W rozdziale drugim omawiana jest ewolucja świadomego umysłu. Działanie ewolucji pokazane zostaje w skali gatunków obdarzonych umysłami oraz w skali pojedynczego mózgu. Uwaga zwrócona jest na podobieństwa struktur organicznych i struktur w maszynach cyfrowych, co jest podstawą do mówienia o naturalnych automatach. Pokazane zostają cztery rodzaje istot, które wyodrębnia Dennett. Są to istoty darwinowskie, skinnerowskie, popperowskie i gregoriańskie. Przedstawiony zostaje sposób, w jaki Dennett łączy prawo efektu z procesem doboru naturalnego. Służy to opisaniu sposobu, w jaki z prostych organizmów żywych wyewoluowały istoty obdarzone świadomymi umysłami. Świadomość ich określona jest jako kompleks memów kultury.

Rozdział trzeci dotyczy funkcjonalistycznej teorii umysłu, zachowującej architekturę równoległą organicznego mózgu. Punktem wyjścia jest tutaj krytyka dualizmu i wprowadzenie w jego miejsce modelu szkiców złożonych. Przedstawione są założenia proponowanej przez Dennetta teorii świadomości. Jest ona omawiana ze zwróceniem uwagi na neurofizjologiczne struktury mózgu, z jednej strony, oraz podzielone na moduły schematy systemów funkcjonalnych, które stosuje psychologia poznawcza i sztuczna inteligencja, z drugiej. Pokazana zostaje trudność w odróżnieniu zafałszowania wspomnień od zafałszowania bezpośrednich doznań. Przedstawione są również pewne uściślenia terminologiczne dotyczące świadomości oraz zagadnienia sprawiające teorii funkcjonalistycznej najwięcej kłopotu, dotyczące jaźni i wyobrażeń.

W rozdziale czwartym omówione jest zagadnienia odwzorowania struktur językowych w mózgu oraz możliwość stworzenia artefaktów, które będą dokładnymi kopiami tego, co wytworzył trwający tysiące lat proces ewolucji na drodze doboru naturalnego. Podjęte zostają problemy języka wewnętrznego oraz możliwości zaistnienia inteligencji. Przedstawione są warunki, które powinna, zdaniem Dennetta, spełniać hipoteza zapisu mózgowego. Omawiane są różnice pomiędzy tradycyjnymi modelami stosowanymi przez nauki poznawcze a sieciami połączeń proponowanymi przez koneksjonizm. Poruszony zostaje też problemu bólu i możliwości odczuwania go przez maszynę. Rozdział kończy się opisem badań, prowadzonych w laboratorium sztucznej inteligencji w MIT.

Kategoria umysłu pojawiła się w filozofii wcześnie. Już joński filozof przyrody Anaksagoras z Kladzomen różnił się od swych poprzedników tym, że uważał umysł [*nous*] za substancję wchodzącą w skład rzeczy żywych, co odróżnia je od materii nieożywionej. Zdaniem Arystotelesa wprowadził on tylko umysł do filozofii i nie zrobił z niego większego użytku. Arystoteles podał złożoną doktrynę, w której odróżnił umysł od duszy. Śmiertelną duszę posiadały wszystkie rośliny i zwierzęta, a nieśmiertelny i bezczasowy umysł tylko wyróżniona wśród nich mniejszość, zajmującą się wyższymi funkcjami myślenia abstrakcyjnego.

Kartezjusz jest autorem bardzo wpływowej koncepcji, nazywanej od jego nazwiska dualizmem kartezjańskim. Zgodnie z tą koncepcją umysły i ciała są substancjami różnego rodzaju. Przedmioty materialne są czasoprzestrzenne, a przedmioty mentalne tylko czasowe.

Istnieją również pomiędzy nimi liczne różnice jakościowe. Jakość świadomego doświadczania bólu jest nieporównywalna z jakościami przedmiotów materialnych. Wyodrębnienie dwóch różnych substancji spowodowało jednak problem interakcji pomiędzy nimi.

Jedną z prób rozwiązania tego problemu był proponowany, między innymi przez Gottfrieda Wilhelma Leibniza, paralelizm, w którym substancja materialna i mentalna nie oddziałują na siebie przyczynowo. Zdarzenia zachodzą w nich równolegle. Wywołało to z kolei problem zapewnienia zgodności pomiędzy przebiegiem zdarzeń fizykalnych i mentalnych. Okazjonalizm proponował jako sprawcę tej zgodności, pośredniczącego w relacji przyczynowo – skutkowej Boga. Problemu wzajemnych oddziaływań umysłu i materii unikał również idealizm, którego najbardziej znanym przedstawicielem jest George Berkeley. Uważał on świat za złożony wyłącznie z umysłów i ich treści, a więc również materia była fenomenem mentalnym. Epifenomenalizm natomiast proponuje rozwiązanie, w którym fenomeny mentalne są skutkami lub produktami ubocznymi złożonych systemów fizykalnych. Zachowuje on jednostronne przyczynowanie zdarzeń mentalnych zdarzeniami fizykalnymi. Nie rozwiązuje to problemu możliwości oddziaływań pomiędzy dwiema odmiennymi substancjami mózgu i umysłu.

Unika tego problemu materializm, który uznaje tylko jeden rodzaj substancji. Zgodnie z nim zarówno umysły jak i ciała składają się z substancji materialnej. Charakterystyki neurofizjologiczne mózgu i introspekcyjne opisy stanów psychologicznych są jednak różne. Różny jest również sposób dostępu do własnych i cudzych stanów psychologicznych. Potrzebny jest obiektywny sposób ich opisu, za jaki uważano behawioryzm. W przypadku behawioryzmu nie można jednak stwierdzić, czy za każdym razem opisuje on taki sam, bo przecież dostępny tylko introspekcyjnie przedmiot. Teoria identyczności natomiast, która sprowadza zdarzenia mentalne do zdarzeń neuronalnych staje przed problemem redukcji. W przestrzeni sieci neuronalnej nie ma miejsca na przekonania, intencje, ból, czy rozumienie.

Bliską materializmowi koncepcją jest funkcjonalizm. Na ogół zakłada on jeden rodzaj substancji, która podlega prawom fizykalnym, a procesy mentalne opisuje językiem opisu stanów maszyn komputacyjnych. Nie jest istotne podłoże [*hardware*] tych procesów, a tylko ich realizacje i spełniane przez nie funkcje. Funkcjonalizm wywołuje problem wystarczalności podłoża procesów komputacyjnych, a w szczególności możliwości zrealizowania świadomego umysłu na jakimkolwiek innym podłożu, niż podłoże organicznego mózgu.

Przedstawicielem reprezentacjonistycznej koncepcji umysłu jest Jerry Fodor. Postuluje on system symboli spełniających funkcje reprezentacji mentalnych. Składają się one na ustalony biologicznie język myśli, analogiczny do kodu stosowanego w maszynach komputacyjnych. Wywołuje to jednak problem znaczenia symboli, które w języku myśli zależne są tylko syntaktycznie. Obrazuje to argument chińskiego pokoju podany przez Johna Searle'a. Zdania w języku chińskim, a wiec i zdania w języku myśli mogą być tam produkowane bez znajomości ich znaczenia.

Eliminatywizm, który głoszą Patricia i Paul Churchland głosi, że kategorie intencjonalne takie jak przekonania, dążenia i intencje mogą okazać się nieprzydatne i lepiej jest zastąpić je kategoriami dotyczącymi innego poziomu opisu. Może nim być na przykład poziom neurologiczny. Tego rodzaju redukcjonizm spotyka się z zarzutem mówiącym, że skoro odrzucone zostały przekonania, to nie można mieć również przekonań, co do samego

eliminatywizmu. Zamiast traktować przekonania i inne kategorie intencjonalne jako rzeczywiste cechy świata, można potraktować je tak, jakby były mu przypisane analogicznie do sposobu przypisania kuli ziemskiej długości i szerokości geograficznej. Koncepcje taką John Heil w *Philosophy of Mind* nazywa interpretacjonizmem. Jednym z przedstawicieli jest Donald Dawidson. Jego interpretacjonizm skupia się jednak na postawach propozycjonalnych, co różni go od koncepcji Daniela Dennetta.

Dennett stosuje różne poziomy interpretacji umysłu. Poprzez opis jego poszczególnych, funkcjonalnie wydzielonych modułów, stara się podać projekt, na podstawie którego będzie mógł zostać zrealizowany sztuczny świadomy umysł. Nie został on jak dotychczas zaprojektowany i skonstruowany. Dennett podaje jednak kierunki oraz metody badań dotyczących tego zagadnienia. W badaniach tego rodzaju odrzucone zostają rozwiązania mistyczne, pozostające poza zasięgiem możliwości poznawczych człowieka i wykorzystany zostaje dorobek nauk szczegółowych. Dennett nawiązuje do nich, począwszy od biologii i teorii ewolucji, a skończywszy na naukach komputerowych i cybernetyce. Pokazuje świadomy ludzki umysł jako przedmiot najbardziej ze wszystkich przedmiotów we wszechświecie złożony, który jest jednak częścią tego wszechświata i podlega tym samym co on prawom. Możliwe jest więc w zasadzie poznanie umysłu i skonstruowanie go jako artefaktu. Nawet jeżeli artefakt będący dokładna repliką ludzkiego umysłu miałby nigdy nie powstać, to sztuczne oko lub ucho już znalazły zastosowanie. Pewne inne moduły umysłu, jak na przykład moduł pamięci, są opracowywane i mogą wkrótce zostać wykorzystane.

1. Intencjonalność

1.1. Intencjonalność a referencyjność. Tezy F. Brentano, kryteria R. Chisholma

Referowanie poglądów Daniela Dennetta należy zacząć od pokazania, w jaki sposób odnosi się on do tradycji filozoficznej i jakich narzędzi filozoficznych używa w swej koncepcji. Referencyjność jest istotna dla intencjonalności, gdyż założenie mówiące, że język dotyczący zdarzeń mentalnych nie może zawierać terminów referencyjnych, pozwala uniknąć potraktowania myśli, pojęć, przedstawień i innych treści umysłu jako zdarzeń fizykalnych. Referencyjność, definiowana jako relacja pomiędzy nazwą a przedmiotem nazywanym, nie zawsze jest możliwa do ustalenia. Dla odróżnienia terminów referencyjnych od niereferencyjnych Dennett proponuje wyróżnienie istnienia w słabym i w silnym sensie. W sensie słabym istnieją jednostki miary, (takie jak mila, metr, czy stopień). Tu należy również wzgląd na coś, wydany lub usłyszany przez kogoś głos. W silnym sensie istnieją przedmioty konkretne, takie jak np. stół. Różnica pomiędzy tymi dwoma rodzajami istnienia polega na trudności we wskazaniu przedmiotu, w przypadku istnienia w sensie słabym, przez co nie można określić relacji referencji. Jeżeli rozważamy głos, to nie wiadomo, czy należy wskazać na drgające cząsteczki powietrza, struny głosowe mówiącego, fale radiowe, czy też zapis głosu na płycie. Ontologiczny status głosu nie jest do końca jasny, przez co „głos" należy uznać za termin niereferencyjny.

Założenie niereferencyjności języka opisującego zdarzenia mentalne pozwala uniknąć zawarcia w założeniach teorii, mających pojawić się dopiero w jej konsekwencji, wniosków. Docieranie do nich rozpoczyna Dennett stawiając pytanie: „Czym są zdarzenia mentalne?"[1]

[1] D.C. Dennett, *Brainstorms*, s. XV.

Odpowiada na nie, że: „Każde zdarzenie mentalne jest (identyczne z) pewnym zdarzeniem fizykalnym w mózgu.” Jednak na pytanie: „Co wspólnego mają dwa stworzenia, kiedy oba są przekonane, że śnieg jest biały...?” nie udziela on odpowiedzi, że „W każdym przypadku, kiedy wspólne jest coś mentalnego, wspólne jest również coś fizykalnego – np. mózgi tych stworzeń są w tym samym stanie fizykalnym lub mają takie same cechy fizykalne.” Nie uznaje on więc teorii identyczności typów zdarzeń mentalnych w umysłach z typami zdarzeń fizykalnych w mózgach. Uznaje natomiast teorię identyczności danego egzemplarza zdarzenia fizykalnego w mózgu z danym egzemplarzem zdarzenia mentalnego w umyśle.

Zgodnie z teorią identyczności typów dla każdego mentalnego predykatu M, ze słownika intencjonalnego, istnieje pewien fizykalny predykat P, ze słownika nauk fizykalnych. Dla pewnego podmiotu x można to sformalizować:

$$(x)\ (\mathrm{M}x \equiv \mathrm{P}x)$$

Jest to formuła redukcjonizmu zastępującego predykaty mentalne predykatami fizykalnymi o tej samej ekstensji. Uważane jest to obecnie za zbyt silne wymaganie i Dennett proponuje w zamian, umieszczenie po prawej stronie równoważności funkcji i stanów funkcjonalnych maszyny Turinga[2]. Wtedy:

(*x*) (M*x* ≡ *x* jest realizacją pewnej maszyny Turinga *k* w stanie logicznym A)

Dwa podmioty mające te same przekonania nie muszą być w takich samych stanach fizykalnych, lecz muszą być w identycznych stanach funkcjonalnych określonych logicznie dla maszyny Turinga. Jest to jednak ciągle redukcjonizm stanów mentalnych do stanów funkcjonalnych wyrażonych w języku teorii automatów. Nie ma podstaw, by przypuszczać, że różne systemy intencjonalne wykorzystują ten sam program działania, więc Dennett proponuje rozwiązanie kolejne:

(*x*) (*x* jest przekonany, że *p* ≡ można przewidzieć przypisanie *x* – owi przekonania, że *p*)

Wydaje się to zwykłym powtórzeniem treści, ale będzie inaczej, jeżeli prawa strona usystematyzowana zostanie w sposób podobny do tego, w jaki opisywane są stany funkcjonalne maszyny Turinga. Opis automatów komputacyjnych nie redukuje ich, a tylko systematyzuje ich stany funkcjonalne i taką metodą proponuje Dennett dla usystematyzowania stanów posiadania przekonań przez systemy intencjonalne.

Przedstawienie swych poglądów w *Content and Consciousness* Dennett rozpoczyna od ulokowania ich w tradycji brentanowskiej[3]. W pracy *Psychologie vom empirischen Standpunkte*, z 1874 r., Brentano sformułował rozróżnienie pomiędzy fenomenami mentalnymi a fizykalnymi. Zgodnie z nim fenomeny mentalne są intencjonalne, w rozumieniu intencjonalności przejętym z filozofii scholastycznej. Intencjonalność ta oznacza nakierowanie na przedmiot konkretny lub też tylko umysłowy. Nie można chcieć, przedstawiać sobie lub twierdzić bez przedmiotu tego chcenia, przedstawienia lub twierdzenia. Istnienie takiego przedmiotu jest istnieniem szczególnego rodzaju i zostało przez Brentano nazwane istnieniem intencjonalnym.

[2] Por. D.C. Dennett, *The Intentional Stance*, s. 66.
[3] Por. D.C. Dennett, *Content and Consciousness*, s. 19-27.

Użycie zwrotów mówiących o nakierowaniu podmiotu bezpośrednio na przedmiot intencjonalny, tak jak to ma miejsce w języku dotyczącym przedmiotów konkretnych, wprowadza, zdaniem Dennetta, nieporozumienia ontologiczne i powinno zostać zastąpione użyciem zwrotów odnoszących podmiot do przedmiotu intencjonalnego poprzez treść, znaczenie lub twierdzenie. Intencjonalność w takim ujęciu, nie służy odróżnieniu fenomenów mentalnych od fizykalnych, lecz podziałowi zdań na zdania intencjonalne i fizykalne.

Dla ujednolicenia języka intencjonalnego i wyróżnienia zdań intencjonalnych lub użytych intencjonalnie, Dennett proponuje zastosowanie trzech kryteriów Chisholma[4]. Dzięki nim nie zakładamy, że są jakiekolwiek zaktualizowane fenomeny (myśli, przekonania, pragnienia), których dotyczy zdanie intencjonalne. Brzmią one następująco:

1. Dane proste zdanie orzekające jest intencjonalne, jeżeli używa wyrażenia rzeczownikowego – nazwy lub opisu – w taki sposób, że ani to zdanie, ani jego zaprzeczenie nie implikuje, że jest cokolwiek, do czego zdanie to się odnosi.
2. Każde jednoznaczne zdanie, które zawiera klasę propozycjonalną jest intencjonalne, jeżeli ani ono, ani jego zaprzeczenie nie implikuje, że ta klasa propozycjonalna jest prawdziwa lub fałszywa.
3. Jeżeli A i B są dwoma nazwami lub opisami opisującymi tę samą rzecz lub rzeczy i zdanie P różni się od zdania Q tylko tym, że zawiera A, podczas gdy Q zawiera B, to P i Q są intencjonalne jeżeli prawdziwość jednego z nich wraz z prawdziwością A i B, które są współdesygnatami, nie implikuje prawdziwości drugiego z nich.

Pierwsze kryterium sformułowane jest w taki sposób, że przedmiot intencjonalny nie musi istnieć. W drugim z kryteriów, niemożliwość określenia prawdy lub fałszu klasy propozycjonalnej powstrzymuje od przesądzenia o istnieniu lub nieistnieniu jej przedmiotu. Konstrukcję trzeciego z kryteriów zobrazować można za pomocą przykładu Quine'a, wykorzystującego współdesygnowanie przez nazwy własne tej samej osoby, (Cyceron i Tuliusz)[5]. Jeżeli ktoś nie wie, że Cyceron i Tuliusz to jedna i ta sama osoba, to zadenuncjonowanie Katyliny przez Cycerona nie pociąga za sobą zadenuncjonowania go przez Tuliusza. Zdanie wyrażające przekonanie o denuncjacji jest więc zdaniem intencjonalnym.

Kryteria Chisholma pomagają uporządkować dualistyczne zróżnicowanie przedmiotów i treści. Pierwsze z nich odnosi się do przedmiotów, natomiast drugie i trzecie do treści. Jeżeli okaże się możliwym zastąpienie zdań przedmiotowych zdaniami wyrażającymi postawy propozycjonalne, to o intencjonalności mówić będzie można przy użyciu samej tylko referencji i prawdziwości klas propozycjonalnych. Dla niektórych zdań, takich jak np. „Wierzę w smoki.", jest to łatwe, gdyż zdanie takie zastąpić można zdaniem: „Wierzę, że smoki są." Nawet, jeżeli nie wszystkie zdania przedmiotowe zastąpić można zdaniami wyrażającymi postawy propozycjonalne z taką łatwością, to sama możliwość pokazuje znaczenie treści dla odróżnienia zdań intencjonalnych od zdań nieintencjonalnych. Ograniczając wypowiedź do wyrażenia postawy propozycjonalnej nie implikujemy konieczności wskazania przedmiotu i sprowadzamy problem na poziom opisu. Nie jesteśmy już nastawieni na przedmiot, a tylko na mówienie o przedmiocie. Wielu autorów rozumie brentanowskie rozróżnienie, jako przypisanie fenomenom mentalnym treści lub znaczenia, w odróżnieniu od fenomenów fizykalnych. Bycie indywidualnym fenomenem jest wtedy kwestią specyficznego opisu lub odnoszącego się do tego fenomenu twierdzenia.

[4] Tamże, s. 22.
[5] Por. W.V.O. Quine, *Słowo i przedmiot*, s. 169-170.

Zdania intencjonalne są intensjonalne. Można przyjąć, że ekstensja terminu jest klasą wszystkich rzeczy, co do których termin użyty jest prawdziwie. Smok i sfinks mają tę samą ekstensję, którą jest klasa zerowa, ale różne są ich intensje. Logika klasyczna nie uwzględnia rozróżnień intensjonalnych w tym sensie, że wzajemne podstawianie terminów o tej samej ekstensji, bez względu na ich intensje, nie narusza wartości prawdziwościowej zdania. Ponadto zdanie logiczne zawsze jest funkcją prawdziwościową swoich składników, a zastosowanie podanych przez Chisholma kryteriów (2) i (3) pokazuje, że zdania intencjonalne nie podlegają tej regule, więc fenomeny intencjonalne są nieredukowalne do fenomenów fizykalnych, do opisu których stosuje się reguły zdań logicznych.

1.2. Postawy propozycjonalne w systemach intencjonalnych

Rozważając przekonania, dążenia, myśli lub odczucia **systemu intencjonalnego** można uznać je za **postawy propozycjonalne**. Zdaniem Dennetta nieuniknionym wydaje się potraktowanie ich jako stanów psychologicznych. Nie można jednak wtedy zapożyczać terminu **postawy propozycjonalne** [*propositional attidudes*] od Russella, gdyż u niego nie są one psychologiczne. Zgodność z sensem tego terminu wymagałaby spełnienia pewnych warunków.

Dopuszczalne mają być trzy stopnie dowolności w formule postawy propozycjonalnej. Są to: (1) stopień dotyczący osoby, przykładowo (*x* jest przekonany, że *p* lub *y* jest przekonany, że *p*); (2) stopień typu postawy, czyli (*x* jest przekonany, że *p* lub *x* boi się, że *p* lub *x* ma nadzieję, że *p*); oraz (3) stopień sądu logicznego, tzn. (*x* jest przekonany, że *p* lub *x* jest przekonany, że *q*). Jeżeli chcemy charakteryzować stany psychologiczne danej osoby za pomocą postaw propozycjonalnych, to musimy podać odpowiednie dla tych postaw sądy logiczne. Wymaga to od nas określenia tego, co należy uznać za dwa różne sądy logiczne, a co za jeden. W literaturze dotyczącej tematu spotykamy trzy odmienne określenia sądów logicznych[6]:

1. Sądy logiczne są zdanio-podobnymi całościami zbudowanymi zgodnie z syntaksą. Sądy, tak jak zdania, dopuszczają rozróżnienie **typ – egzemplarz** [*type – token*]. Różne sądy, należące do tego samego typu, znajdują się w umysłach (lub mózgach) podmiotów mających te same przekonania. Do tego właśnie poglądu odwołują się, zdaniem Dennetta, przedstawiciele nauk poznawczych, w debacie dotyczącej form reprezentacji mentalnych. Zadają oni pytanie: „Czy wszystkie reprezentacje mentalne są zdaniowe, czy też niektóre są wyobrażeniowe lub analogowe?”
2. Sądy logiczne są zbiorami [*sets*] światów możliwych. Dwa zdania wyrażają ten sam sąd, jeżeli są prawdziwe dla dokładnie tego samego zbioru światów możliwych. Pogląd ten nie zakłada własności syntaktycznych sądów. Nie można też mówić o nich jako o mających swe egzemplarze [*tokens*] w mózgu, umyśle lub w postaci słowa pisanego.
3. Sądy logiczne są czymś takim, jak zbiory lub układy przedmiotów w świecie. Korespondencyjna teoria prawdy głosi, że tym, co czyni zdanie prawdziwym, jest korespondencja z faktami w świecie, gdzie fakt okazuje się być gwarantem prawdziwości sądu logicznego.

[6] Por. D.C. Dennett, *The Intentional Stance*, s. 121.

Dennett uważa, że te grupy poglądów łączy zbiór trzech innych wymagań dotyczących poprawności użycia sądów logicznych w teorii. Pochodzą one od Fregego, którego pojęcie *myśli* jest podstawą współczesnego rozumienia sądów. Zgodnie z tym poglądem *myśl* musi mieć trzy ściśle określone cechy:

a. Bycie nośnikiem, ostatecznej i dalej niewyprowadzalnej, wartości prawdziwościowej. Jeżeli *p* jest prawdziwe, a *q* fałszywe, to *p* i *q* nie są tymi samymi myślami.
b. Złożoność z intensji, gdzie intensje są rozumiane á la Carnap, jako determinanty ekstensji. Różne intensje mogą determinować tę samą ekstensję, (np. 3^2 lub liczba planet), ale różne ekstensje nie mogą być determinowane przez jedną intensję.
c. Ujmowalność przez umysł.

Frege nie wyjaśnia dalej „ujmowalności” myśli lub sądu logicznego. Pociągałoby to za sobą konieczność zarysowania relacji pomiędzy abstraktami, (np. prawda i fałsz), a umysłem (lub mózgiem), co wiąże się z wieloma trudnościami metafizycznymi. Można by ich jednak uniknąć traktując predykaty postaw propozycjonalnych, tak jak predykaty miar fizycznych. Trzeba by znaleźć analogię pomiędzy, np. „Jest przekonany, że *p*” a „Ma długość *n* metrów”. Przekonanie byłoby tym samym, co znajdowanie się w pewnej określonej relacji do danej wielkości. Chociaż wyjaśnienie takie nie rozwiązuje metafizycznego problemu sądów, to, przez przyrównanie ich do liczb występujących w wynikach pomiarów fizycznych, usuwa problem istnienia jakiegoś specjalnego rodzaju przedmiotów abstrakcyjnych w psychologii.

Psychologia przypisuje sądom rolę przyczynową w zdarzeniach psychicznych. Można więc pytać o funkcję sądów w myśleniu, ale tylko jeżeli będą one konkretami lub będą miały konkretne egzemplarze [*tokens*]. Prowadzi to prosto do wersji (1) rozumienia sądów mówiącej, że sąd logiczny jest całością zdaniopodobną. Nie można przypuszczać, że światy możliwe lub układy przedmiotów i ich własności są „w głowie”, a tylko coś, co jest „w głowie” może pełnić bezpośrednią rolę przyczynową w psychologii.

Można również rozważać sądy za pomocą analogii z liczbami. Relacje formalne pomiędzy sądami zostaną wtedy użyte do utworzenia nauki zajmującej się „pomiarem” stanów psychologicznych, w ten sam sposób, w jaki użyte są relacje pomiędzy liczbami w fizyce. Otwartym pozostaje pytanie, czy w istocie przejawiającej postawę propozycjonalną coś odzwierciedla klasę wyznaczoną danym sądem. Należałoby znaleźć wewnętrzną strukturę podmiotu mającego przekonania, które można potraktować jako system zdań lub system wewnętrznych reprezentacji ujmowanych przez umysł na zasadzie podobnej do ujmowania liczb w procesach fizycznych.

Putnam i eksternaliści argumentują jednak w sprawie treści, że to, co ujmowalne dla umysłu nie może jednocześnie determinować ekstensji lub wartości prawdziwościowej. Przemawia za tym eksperyment myślowy Putnama z Bliźniaczą Ziemią[7]. Planeta ta jest prawie dokładna repliką Ziemi, włącznie ze wszystkimi poszczególnymi ludźmi, rzeczami i zdarzeniami. Jedyna różnicą jest to, że rzeki, jeziora i inne zbiorniki wodne nie zawierają H_2O, lecz XYZ, które jest chemicznie różne, lecz zwykłymi metodami nieodróżnialne od H_2O. Mieszkańcy Bliźniaczej Ziemi nazywają ten płyn wodą. Są oni fizykalnymi, jak również psychologicznymi replikami mieszkańców Ziemi, więc wyznają te same teorie na poziomie,

[7] Por. H. Putnam, *Znaczenie wyrazu „znaczenie”*, w: *Wiele twarzy realizmu i inne eseje*, s. 93-184.

na którym H_2O i XYZ są nieodróżnialne. Kiedy jednak mieszkańcy Ziemi mają jakieś przekonania na temat H_2O, przekonania ich replik z Bliźniaczej Ziemi odnoszą się do XYZ. Są to więc przekonania odmienne, gdyż w zależności od tego, czy przedmiotem ich jest H_2O, czy XYZ, będą one prawdziwe dla Ziemian i fałszywe dla ich replik z Ziemi Bliźniaczej lub odwrotnie. Zachodzi tu identyczność stanów psychologicznych, lecz nie ma identyczności twierdzeń. Putnam wyciąga z tego wniosek, że znaczenie nie znajduje się „w głowie" lub, że znaczenie nie determinuje ekstensji. Eksperyment Putnama bazuje na koncepcji rodzajów naturalnych i **sztywnego desygnowania** [*rigid designation*]. Jednak myśli lub przekonania danego podmiotu zawierają na ogół pewien **funktor indeksowania** [*indexical functor*], taki jak: teraz, ten, tutaj. Trudno przypuszczać, że stany wewnętrzne zależne będą od czegoś całkowicie zewnętrznego, przez żaden z takich funktorów niewskazanego, a czymś takim jest struktura chemiczna wody, zarówno H_2O jak i XYZ.

Dennett interpretuje to tak, że sąd logiczny może zostać ujęty przez umysł, jeżeli ujęte zostaje konkretne, wyrażające ten sąd zdanie. Zdanie jako takie, nie musi być nośnikiem wartości prawdziwościowej, ani jego intensja nie musi determinować ekstensji. W skład zdania wchodzą symbole lub reprezentacje przedmiotów rzeczywistych pozostające w specyficznych dla nich zależnościach syntaktycznych języka **mentalnego** [*Mentalese*]. Dennett stwierdza, że nie można jednak w pełni zastosować do sądów funktorów indeksowania. Egzemplarz zdania danego typu może wyrażać różne twierdzenia w zależności od kontekstu, od innego zdania, w którym został użyty funktor indeksowania lub od innej sytuacji opisanej określonymi zdaniami wyrażającymi jej kontekst. Zdania rozważane bez przedmiotów, do których się odnoszą nie wystarczają dla zdeterminowania przekonań podmiotu w sensie jego postaw propozycjonalnych. Podany przez Putnama przykład świadczy o tym, że stany wewnętrzne mieszkańca Ziemi i jego odpowiednika z Ziemi Bliźniaczej, w postaci zdań języka mentalnego, który ma formę mózgowego zapisu, mogą być identyczne, podczas gdy znaczyć będą one coś innego. W przypadku tym, aby poznać znaczenie, trzeba dotrzeć do struktury chemicznej wody, H_2O lub XYZ. Jeżeli nie jest to możliwe, nie można mówić o sądach logicznych.

Tam, gdzie nie można zastosować postawy propozycjonalnej Dennett zaproponował zastosowanie **postawy zdaniowej** [*sentential attitude*][8]. Próba określenia stanów psychologicznych za pomocą postaw zdaniowych również napotyka na wiele problemów związanych z tym, że oparta jest ona na syntaktyce. Bliźniacy psychologiczni lub maszyny Turinga mogą określić swe syntaktyczne podobieństwo na pewnym poziomie. Jednak dla pewnych stanów, którym odpowiadają określone poziomy opisu, takie same stany na wyjściu związane będą z odmienną syntaktyką. Pozycją pośrednią pomiędzy postawą propozycjonalną, a postawą zdaniową jest przyjęcie **psychologii postawy pojęciowej** [*notional attitude psychology*]. Mieszkaniec Ziemi i jego replika z Ziemi Bliźniaczej mają dokładnie takie same **postawy pojęciowe** [*notional attitudes*], a wszelkie różnice w ich postawach propozycjonalnych wynikają z różnic środowiska, w którym żyją.

Mieszkańcy Ziemi żyją w jednym i tym samym świecie rzeczywistym, lecz w różnych światach pojęciowych, chociaż w pewnych zakresach światy te mogą się pokrywać. Świat pojęciowy powinien być pojmowany jako pewien świat fikcyjny wypełniony przedmiotami pojęciowymi, co do których podmiot ma jakieś przekonania. Odrzucając solipsyzm metodologiczny trzeba zauważyć, że pewne przedmioty świata rzeczywistego „pasują" do podmiotowego świata pojęciowego. Określenie relacji pomiędzy tymi dwoma rodzajami

[8] Por. D.C. Dennett, *Intentional Stance*, s. 130-150.

przedmiotów nasuwa trudności, co często jest przyczyną wycofania się na stanowisko solipsyzmu metodologicznego, który Dennett rozumie jako wersję zaproponowanego przez Husserla époché lub wzięcia w nawias. Jednak alternatywą dla postaw propozycjonalnych (semantyka) i zdaniowych (syntaktyka) nie do końca może być fenomenologia. Różnica pomiędzy fenomenologią a rozwiązaniem proponowanym przez Dennetta polega na tym, że, podczas gdy fenomenologia stosuje redukcję fenomenologiczną opartą na introspekcji, Dennett chce określać świat pojęciowy danego podmiotu z zewnątrz. To, co uprawiali Brentano i Husserl określa on jako **autofenomenologię**. Proponowane przez siebie podejście nazywa **heterofenomenologią**[9].

Zamiast perspektywy podmiotu w pierwszej osobie, którą stosował Kartezjusz, Dennett wprowadza heterofenomenologiczną perspektywę podmiotu w osobie trzeciej. Chce w ten sposób uniknąć problemu stwierdzalności istnienia świadomości w podmiocie. Nie ma sposobu na zbadanie cudzej świadomości od wewnątrz, a więc otaczający nas ludzie mogą być tylko *zombi* – istotami bez świadomości, jak też świadomość posiadać może cokolwiek, czego obserwowane z zewnątrz zachowanie potraktować można jako wynik świadomych decyzji. Postępowanie takie nazywa Dennett przyjęciem **nastawienia intencjonalnego** [*intentional stance*]. Świadomość, jako fenomen dostępny tylko na drodze introspekcji, nie interesuje Dennetta.

Umysły zależne są od zewnętrznego, oddziałującego na nie zdarzeniami zewnętrznymi, środowiska. W nich samych natomiast zachodzą zdarzenia wewnętrzne, które również stanowią rodzaj środowiska. Problemem jest znalezienie relacji pomiędzy środowiskiem zewnętrznym a wewnętrznym. Nie jest to jednak konieczne dla semantyki odnajdywanej w środowisku wewnętrznym. Potrzebny jest tu model w rodzaju semantyki Tarskiego. Dennett zauważa, że pomysł światów pojęciowych jest pomysłem na model, niekoniecznie rzeczywisty lub prawdziwy model reprezentacji mentalnych. „Nie składa się on z reprezentacji, lecz z tego, co reprezentowane"[10] – pisze Dennett i chodzi mu o to, że jest to świat, w którym podmiot żyje, a nie świat, który żyje w podmiocie. Tak, zdaniem Dennetta należy rozumieć Brentano.

Świat fikcyjny nie istnieje w zwykłym sensie. Dennett traktuje jego zawartość jako referencyjne odniesienie dla podmiotowej reprezentacji. Świat ten składa się z przedmiotów intencjonalnych i może być również rozważany formalnie, jak to robią Hintikka lub Talnaker. Psychologia postaw pojęciowych może być sposobem na uniknięcie problemów semantycznego odniesienia, występujących w przypadku postawy propozycjonalnej i problemu braku odniesienia postawy zdaniowej. Strategia taka wykorzystywana jest w eksperymentach z zakresu sztucznej inteligencji, gdzie do symulacji środowiska używa się komputera. Takie symulowane środowisko jest odpowiednikiem świata pojęciowego.

Dennett proponuje założenie mówiące, że w postawach propozycjonalnych mamy do czynienia z przekonaniami **de re**. Pisze następnie, że zamiast przekonań **de re**, odpowiednich dla postaw propozycjonalnych, rozważać można przekonania **de dicto**, stosowne dla postaw pojęciowych, które dopiero w połączeniu ze środowiskiem zewnętrznym podmiotu, będą wystarczające dla określenia jego postaw propozycjonalnych.

---

[9] Por. D.C. Dennett, *Consciousness Explained*, s. 72.
[10] D.C. Dennett, *Intentional Stance*, s. 154.

1.3. Krytyka behawioryzmu z punktu widzenia intencjonalności

Behawioryści argumentują, że jeżeli idiomy intencjonalne nie dają się uzgodnić z językiem opisu fenomenów fizykalnych, czyli z językiem naukowym, to należy ich unikać. Chcą oni odkryć ekstensjonalne prawa rządzące zachowaniem ludzi i zwierząt, a idiomy intencjonalne sprowadzić do poziomu języka naturalnego. Dennett krytykuje takie podejście i pokazuje, że całkowite zastąpienie zdań intencjonalnych zdaniami języka nauki dotyczącej świata fizykalnego jest niemożliwe. Wyjaśniając zachowania zwierzęcia nie sposób ograniczać się wyłącznie do języka opisującego poruszenia mięśni i układu kostnego. Zwierzę „wie", „ma przekonanie" lub „świadomość" tego, gdzie jest pożywienie lub wyjście z klatki.

Takie właśnie sposoby opisu zachowania się ludzi i zwierząt są powszechnie stosowane. Nie są one oparte na wyjaśnieniu przyczynowym, w sensie terminu zbliżonym do Hume'owskiego, w którym przyczyna określana jest niezależnie od skutku. W przeciwnym wypadku stwierdzanie przyczyny i skutku nie jest empirycznie prawdziwe, lecz analityczne, prawdziwe tylko ze względu na znaczenie słów. Taka zależność jest niemożliwa, jeżeli poprzednikiem jest intencja, a następnikiem działanie, co można powiedzieć o wszystkich związkach poprzednik – następnik scharakteryzowanych intencjonalnie. Dennett rezerwuje pojęcia przyczyny i skutku dla języka opisującego zdarzenia fizykalne. W języku tym charakterystyka przyczyny nie może zawierać niczego, co prowadziłoby do analitycznego wyprowadzenia z niej skutku.

W wyjaśnieniach intencjonalnych spotykamy sytuację przeciwną. Niemożliwa jest charakterystyka poprzednika niezależnie od następnika. Jeżeli pewien scharakteryzowany intencjonalnie poprzednik zastąpiony zostanie innym, oznacza to ekstensjonalną parafrazę, co już na wstępie zostało odrzucone. Można założyć, że opisywany fenomen posiada inną intencjonalną charakterystykę, ale będzie to niezgodne z podstawową zasadą intencjonalności, dla której inna charakterystyka oznacza inny fenomen.

Jeżeli podmiot zamierza wykonać jakieś działanie wynikające z jego intencji, wciąż może on wykonać to samo działanie pod wpływem przyczyny z tą intencją niezwiązanej. Wtedy jednak, aby jego działanie scharakteryzować, trzeba posłużyć się wyjaśnieniem przyczynowo – skutkowym. Postępowanie takie prowadzi zawsze do użycia w nim stanu wewnętrznego podmiotu, co jest równoznaczne z przejściem na poziom charakterystyki intencjonalnej. Nie jest ona poszukiwaniem reguł i mechanizmów, jak ma to miejsce w wyjaśnieniach behawioralnych, Wyjaśnienia intencjonalne nie podlegają regule mechanicznego łączenia zdarzenia, opisywanego jako poprzednik, ze zdarzeniem, opisywanym jako następnik, co stawia naukę intencjonalną w opozycji do innych nauk.

Nieredukowalność intencjonalności świadczyć może o ograniczeniach psychologii behawioralnej. Zastosowanie alternatywy w postaci całkowicie intencjonalnej psychologii wprowadziłoby ogromne przeobrażenia do wszystkich nauk. Równie dobrze mogą one wciąż funkcjonować równolegle, w niezmienionej formie.

W języku opisującym zachowanie przy użyciu schematu przyczynowo-skutkowego nie można uniknąć zwrotów intencjonalnych, co świadczy o uwikłaniu w ten opis intencjonalności. Ma to szczególne znaczenie w naukach dotyczących organizmów żywych. Argument przemawiający za nieredukowalnością zdań intencjonalnych do ekstensjonalnych odwołuje się do braku behawioralnej podstawy, do której można by odnieść wyrażenie intencjonalne. Ten brak zakończenia intencjonalnych łańcuchów połączeń poprzedników i

następników pozostawia wiele miejsca na zdarzenia wewnętrzne. Dennett zadaje pytanie o ekstensjonalny opis intencjonalnych stanów wewnętrznych i przedstawia alternatywę dla nieredukowalności opisu intencjonalnego.

Uniknięcie dylematu jest równoznaczne ze znalezieniem drogi prowadzącej od ruchu i materii do treści i celu oraz z powrotem. Jeżeli zostałoby to osiągnięte, wtedy można by powiedzieć, że pewien stan fizyczny ma znaczenie lub treść. Komputery „przesyłają" i „przechowują" wiadomości, ale nie mają dostępu do ich treści i znaczenia w wymaganym tu sensie. Organizm intencjonalny dokonuje opisu środowiska takim, jakim je widzi, natomiast komputer przeprowadza operacje w taki sposób, w jaki widział je jego projektant i konstruktor. Komputer jest intencjonalny tylko ze względu na intencjonalność swego twórcy.

Ludzie nie zostali skonstruowani w taki sposób, w jaki sami konstruują komputery. Ich intencjonalność musi, więc mieć inne pochodzenie. Można by wprowadzić hipotezę intencjonalnego Boga, który użyczył nam swej intencjonalności. Chcąc uniknąć tej hipotezy Dennett proponuje poszukiwać źródeł intencjonalności żywych systemów w teorii ewolucji na drodze doboru naturalnego. Opis intencjonalny zakłada, będące efektem doboru naturalnego, dostosowanie do środowiska, gdyż organizm na drodze ewolucji zdobywa wiedzę na temat nakierowania się na odpowiedni przedmiot w jego środowisku. Korzystając z tego Dennett konstruuje teorię centralistyczną, w odróżnieniu o peryferyjnej teorii behawioryzmu, typu bodziec – reakcja. Podczas gdy celem teorii peryferyjnej jest scharakteryzowanie ekstensjonalne zachowań i podporządkowanie ich ogólnym prawom, teoria centralistyczna zaczyna od charakterystyki intencjonalnej. Opisuje ona zdarzenia przy współudziale lub też przy wyłącznym użyciu wyrażeń intencjonalnych. Celem jej jest znalezienie adekwatnych podstaw fizykalnych dla stanów wewnętrznych organizmu, tak, że redukcja zdań intencjonalnych do ekstensjonalnych okaże się możliwa.

1.4. Ewolucja i intencjonalność

Doktryna **pierwotnej intencjonalności** [*original intentionality*] głosi, że, podczas gdy pewne artefakty mają intencjonalność pochodną, zapożyczoną z umysłów ludzkich, ludzie mają **intencjonalność wewnętrzną** [*intrinsic intentionality*], znikąd już dalej nie wyprowadzaną. W takim rozumieniu można ją więc utożsamiać z intencjonalnością pierwotną. Dennett zastanawia się, czy można dokonać takiego utożsamienia. Jest ono poprawne, jeżeli tylko istnieje coś takiego jak intencjonalność pierwotna. Jest on przekonany, że to, co nazywamy intencjonalnością wewnętrzną umysłów ludzkich nie ma związku z intencjonalnością pierwotną i, że nic takiego jak intencjonalność pierwotna nie istnieje. Intencjonalność wewnętrzna umysłów ludzkich jest tylko daleką pochodną procesu ewolucji.

Artefakty mają znaczenie tylko o tyle, o ile my im je nadajemy. Ich intencjonalność jest po prostu wywiedziona z podmiotu intencjonalnego. Komputer nie myśli, ani, sam w sobie, nic nie znaczy. Prawdziwe rozumienie tym różni się od sztucznego rozumienia komputera, że oparte jest o swoją własną intencjonalność. Wielu autorów uważa, że żaden artefakt nie może zostać wyposażony w taki sam rodzaj intencjonalności, jaki mają ludzie. Nie ma takiego komputera lub automatu, który jest autonomicznym myślicielem takim, jakimi są ludzie. Nie ważne jak silną iluzję zdolna jest wytworzyć maszyna.

Dennett rozważa przykład automatu, który rozpoznając pewien sygnał lub przedmiot wprowadza się w stan Q. Czasem dochodzi jednak do mylnej „identyfikacji”, czy też mylnej „percepcji” tego sygnału lub przedmiotu. Wtedy konstruktorzy udoskonalają mechanizm odpowiedzialny za rozpoznawanie. Zdolności rozpoznawcze maszyny mają swe oparcie w intencjonalności jej konstruktorów i do tego miejsca, pisze Dennett[11], filozofowie tacy jak Fodor, Searle, Dretske, Burge i Kripke powinni się z nim zgadzać. Słowo „intencjonalność”, w odniesieniu do automatu, użyte jest tylko metaforycznie. Jeżeli pewien przedmiot lub sygnał zostaje mylnie rozpoznany, należy maszynie użyczyć więcej z ludzkiej intencjonalności, potrzebnej do rozpoznania większej ilości obecnych w przedstawieniu detali. Automat nie ma wewnętrznej intencjonalności, z czym zgadzają się wymienieni filozofowie. Stanowisko Dennetta zaczyna odbiegać od ich poglądów, kiedy stwierdza on, że również umysł ludzki pozbawiony jest tej wewnętrznej intencjonalności. Przyznawanie jej ludziom jest tylko pragmatyczną regułą interpretacji. Umysły ludzkie nie różnią się tak bardzo od automatów. Zawarta w nich i przez nie stosowana intencjonalność wytworzona jest na tej samej zasadzie, co intencjonalność automatu rozpoznającego dany sygnał lub przedmiot. Różnica polega tylko na tym, że nie ma tu konstruktora udoskonalającego mechanizm rozpoznawania, a jest szereg czynników, których działanie Dennett wyjaśnia przy użyciu hipotezy ewolucyjnej.

Posługując się Putnamowskim przykładem Bliźniaczej Ziemi przedstawia problem rozpoznawania przedmiotów nie różniących się zewnętrznym wyglądem, a różnych co do swej wewnętrznej struktury, dostępnej tylko specjalistycznemu badaniu, jako problem niedostatecznie dobrze wykształconego systemu rozpoznawania. Koncepcję intencjonalności pierwotnej uważa on za wątpliwą i co do tego zgadzają się z nim[12] filozofowie tacy jak: Churchlandowie, Davidson, Haugeland, Millikan, Rorty, Stalnaker, Quine, Sellars i Minsky. Przy proponowanym rozumieniu intencjonalności również natura, której działanie oparte jest o proste rozpoznawanie danego sygnału lub przedmiotu, wyposażona jest w intencjonalność. Wszelka inna intencjonalność jest pochodną intencjonalności prostych organizmów żywych, których intencją jest przetrwanie.

Działanie natury można wyobrazić sobie na zasadzie działania automatu, który musi: „poszukiwać”, „rozpoznawać”, „przewidywać”, „unikać”, „wiedzieć” i robić wiele innych intencjonalnych rzeczy. Nie są to jednak czynności rozumiane w sensie, w jakim rozumiane są czynności wykonywane przez ludzi. Działający w taki sposób automat nie jest automatem jedynym i, dla osiągnięcia swych celów, musi współpracować lub współzawodniczyć z innymi automatami. Zwiększa to stopień komplikacji jego zachowań intencjonalnych. Powinien on tworzyć projekty i przewidywać skutki swego działania, podlegać samokontroli i być zdolnym do uczenia się nowych zachowań. Wciąż jednak będzie to zwyczajny automat, zachowujący się zwyczajnie w testach maszyny Turinga. Dennett porównuje ludzi do „maszyn przetrwania” zaprojektowanych tak, by chronić ich geny. Jesteśmy tylko artefaktami a jeżeli posłużyć się porównaniem do Nieporuszonego Poruszyciela Arystotelesa, to nasze geny są niepomyślanymi myślicielami [*Unmeant Meaners*][13]. Nie mają one zdolności myślenia ani świadomości, ale są nośnikami osiągnięć długotrwałego procesu ewolucji na drodze doboru naturalnego.

Problem, z którym się spotykamy używając porównania natury do będącego dziełem określonego twórcy artefaktu, to występujący w przypadku natury brak świadomego

[11] Tamże, s. 291.
[12] Tamże, s. 295.
[13] Tamże, s. 288.

podmiotu aktów twórczych. Wyeliminowanie go z wyjaśnień pochodzenia gatunku ludzkiego jest jednak głównym osiągnięciem teorii ewolucji. Dennett zwraca uwagę, że jej proces interpretować można tak, jakby przejawiał cechy umysłu ludzkiego. Ewolucja dokonuje bardzo subtelnych wyborów, docenia znaczenie wielu złożonych relacji. Kiedy naturalna selekcja „wybiera" robi to z określonego powodu. Proces ten pozbawiony jest jednak jakichkolwiek reprezentacji i reprezentowania.

Pomysł, że jesteśmy tylko artefaktami zaprojektowanymi przez dobór naturalny rzucać może odmienne światło na związek nas, jako podmiotów myślących, z naszymi myślami. Jako artefakty nie nadajemy znaczenia naszym myślom, ani też nie możemy mieć dostępu do będących tym znaczeniem zdarzeń mentalnych. Takie właśnie zarzuty wysuwane są przeciwko przyjęciu wobec procesu ewolucji **nastawienia intencjonalnego**. Dyskutowanie tych zarzutów i obronę swojego stanowiska Dennett rozpoczyna wprowadzając rozróżnienie na **znaczenie naturalne** [*natural meaning*] i **znaczenie funkcjonalne** [*functional meaning*]. Odwołuje się tu do poglądów Dretske'go, chociaż wcześniej o znaczeniu naturalnym mówił Grice.

Znaczenie naturalne (znaczenie$_n$) zdefiniowane jest w sposób, pozwalający uniknąć błędnego reprezentowania. Obejmuje ono całość związków znaczeniowych zdarzeń w jakikolwiek sposób związanych ze zdarzeniem, którego znaczenie jest rozważane. Znaczenie funkcjonalne (znaczenie$_f$) natomiast wykorzystuje sposób wyjaśniania, w którym pewien znak, stan lub zdarzenie może, w jakimś systemie, przy pewnych okolicznościach, reprezentować coś błędnie lub być podstawą orzekania fałszu. Funkcja taka jest funkcją przypisywaną zdarzeniu przez określony podmiot, a znaczenie funkcjonalne związane jest z intencjami, przekonaniami i celami podmiotu, który znaczenie to przypisuje. Dennett przytacza określenie znaczenia funkcjonalnego, które podał Dretske: „Poszukujemy tego, co pewien znak przypuszczalnie *znaczy*, a to, co 'przypuszczalne' jest rozmieniane [*cashed out*] na funkcje tego znaku (lub systemu znaków) we *własnej* ekonomii poznawczej danego organizmu."[14]

W zaproponowanym przez Dennetta, naturalistycznym rozumieniu intencjonalności, zastępujemy własne cele i intencje celami i intencjami natury lub procesu doboru naturalnego. Pytamy następnie, jakie w tym schemacie rozumowania, znaczenie stanów lub sygnałów zostało zaprojektowane, zupełnie jak byśmy pytali o to podmiot świadomy, mający swe własne przekonania mogące powodować błędność interpretacji w przypisywaniu treści. Błędy takie zachodzą w naturalnych artefaktach, czyli w organizmach ludzkich. Problemem jest tutaj wyposażenie biologicznej wersji automatu w zasadę umożliwiającą dokonywanie subtelnych rozróżnień w znaczeniach. Zasada taka powinna zostać wskazana bez sztucznego powodowania **inflacji funkcji naturalnych** organizmów i jednocześnie bez stosowania zbyt **deflacyjnej reguły** rozkładającej wszystkie **znaczenia funkcjonalne** na czyste **znaczenia naturalne** nie dopuszczające błędnego reprezentowania.

Rozważyć można zmylenie systemu percepcyjnego żaby, przez wprowadzenie w jej pole widzenia niewielkiego przedmiotu imitującego muchę. Jeżeli interpretować będziemy sygnał biegnący z oka żaby jako mówiący, że widziana jest mucha, to do systemu percepcyjnego żaby błędną informację wprowadzać będzie oko. Jeżeli jednak zinterpretujemy ten sygnał jako pewną ciemną plamkę przesuwającą się po siatkówce, to oko wprowadzi informacje prawdziwą, a błąd powstanie w części mózgu, do której sygnał dociera później. Możemy zminimalizować nasze interpretacje i wtedy żaba nigdy nie popełni błędu. Rozpoznawanie

[14] Tamże, s. 301.

zawsze dokonywane będzie na poziomie znaczenia naturalnego, do którego doprowadzi nas interpretacja deflacyjna.

Dretske zastanawiał się, czy typy zdarzeń nie są z reguły wyposażane w zbyt dużą porcję znaczenia. Dennett proponuje zmniejszenie tej porcji znaczenia poprzez poddanie funkcji naturalnej pewnej inflacji. Podany zostaje przykład odpowiedzi na pytanie: „Dlaczego tak wiele stworzeń wyposażonych jest w zdolność wykrywania symetrii względem osi pionowej?"[15] Opis deflacyjny sprowadza odpowiedź do stwierdzenia odpowiedniego wzoru na siatkówce oka, co nie daje jednak odpowiedzi na pytanie, dlaczego wyczulenie na ten rodzaj symetrii jest tak istotne.

Zanim powstało wiele artefaktów przejawiających symetrię wokół osi pionowej, w świecie symetrię taką przejawiało zwierzę widziane od przodu. Tym, co mówi symetria pionowa, jest więc: „Ktoś na ciebie patrzy!" Jest to istotna informacja, by rozpoznać zagrożenie w postaci drapieżnika. Zwierzę może również znaleźć się na wprost innego przedstawiciela swojego gatunku i potrzeba wtedy większej czułości przekaźnika informacji. Należy odróżnić impuls identyfikowany z zagrożeniem od impulsu oznaczającego spotkanie w celach pokojowych. Treść takiego przekazu i jej przypisanie to, nie tylko zagadnienie nauk biologicznych, ale i, szeroko pojętej, filozoficznej teorii znaczenia.

W *Machines and the Mental* Dretske głosi, że podstawowa różnica pomiędzy człowiekiem a komputerem polega na tym, że komputery przekazują informacje i dokonują na nich różnorodnych operacji, przy użyciu określonych symboli, ale nie mają dostępu do znaczenia tych symboli, do rzeczy, które są przez nie reprezentowane[16]. Sugeruje to, że Dretske łączy dwie rzeczy: coś ma jakieś znaczenie dla danego organizmu lub systemu oraz ten organizm lub system jest zdolny to rozpoznać w swoim wnętrzu. Nie musi to jednak być znaczenie naturalne. Pominięcie go przenosi zagadnienie na grunt znaczeń poszczególnych funkcji, które, jeśli rozważane są osobno, nie mówią nic o celach i intencjach całego systemu lub organizmu.

Pokazana zostaje w ten sposób możliwość niezależności rozważanego znaczenia od celów i intencji człowieka jako systemu i organizmu. Pominięta jest jednak zależność znaczenia od celów i intencji natury, które są, zdaniem Dennetta, istotne, gdyż przyjęcie ich pozwala pominąć znaczenia wewnętrzne. Prowadzi to do konkluzji mówiącej o niezdeterminowaniu interpretacji znaczeń wyprowadzonych, przez poszczególne podmioty, z intencji i celów natury. Dretske proponuje wprowadzenie zdeterminowania funkcjonalnego na bazie uczenia się. Organizm lub system poddany wielokrotnie działaniu tych samych bodźców może, na bazie uczenia się poprzez asocjacje, wytworzyć pewien określony stan wewnętrzny, który do pewnego stopnia zdeterminuje go funkcjonalnie. W ten sposób znaczenia symboli, którymi posługuje się organizm lub system mogą zbliżyć się do znaczenia naturalnego. Nie mogą go jednak osiągnąć, gdyż aby tego dokonać organizm musiałby wiedzieć wszystko.

Dretske uważa, że nie ma wrodzonych myśli, czy przekonań, wszystkie można wyjaśnić przy użyciu historii życia i zachowań danego organizmu. Dennett nie zgadza się z tym, gdyż zwierzęta przejawiają pewne wrodzone, instynktowne zachowania. Są one tym, co ewolucja lub natura „miała na myśli". Zróżnicowanie proponowane przez Dretske'go jest zróżnicowaniem typu: syntaktyka – znaczenie naturalne a semantyka – znaczenie

[15] Tamże, s. 303.
[16] Tamże, s. 304.

funkcjonalne. Nie wystarcza ono jednak, by pokazać jak stany wewnętrzne systemu lub organizmu nabywają znaczenia w procesie uczenia się.

Za motto rozwiązania proponowanego przez Fodora uznać można zdanie: „Nie ma reprezentacji bez błędnej reprezentacji."[17] W jego reprezentacjonistycznej teorii umysłu tylko stany mentalne mają własności semantyczne. Fodor chce uniknąć problemu rozłączności semantyczno – syntaktycznej, czy też mentalno – fizykalnej i sugeruje, że można to zrobić bez użycia teorii ewolucji. W jej miejsce proponuje on pierwotne [*original*] i wewnętrzne [*intrinsic*] znaczenie naszych reprezentacji mentalnych[18]. Dennett stwierdza, że takie wyjaśnienie jest równie niewystarczające, jak wyjaśnienie Dretske'go. Nie odnosimy w nim znaczenia do przedmiotu zewnętrznego i nie ma żadnej semantyki. Nie zostaje nawiązana żadna relacja organizmu ze środowiskiem zewnętrznym. Organizm nie może więc rozpoznawać błędnie, ani mieć błędnych reprezentacji.

Rozważania stanów mentalnych oparte o możliwość popełnienia błędu lub błędnej reprezentacji prowadzą Dennetta do argumentów Burge'a przeciwko koncepcji indywidualizmu[19], według której indywidualne stany i zdarzenia intencjonalne nie mogą być inne niż są, gdyż indywidualne fizykochemiczne, neuronalne lub funkcjonalne procesy z nimi związane opisywane są nieintencjonalnie, w sposób niezależny od społecznego kontekstu organizmu. Burge argumentuje za fałszywością tezy indywidualizmu posługując się przykładem automatu mylnie rozpoznającego bardzo podobne przedmioty jako takie same. Nic strukturalnie, ani fizykalnie nie zostaje we wnętrzu takiego automatu zmienione, a jednak przypiszemy jego stanom wewnętrznym różne znaczenia, w zależności od tego czy użyjemy go do rozpoznawania jednego, czy też drugiego rodzaju przedmiotu.

W przykładzie tym nie pozostawiono, jak dotąd, miejsca na niezdeterminowanie treści. Zawsze istnieje pewien fakt, do którego, poprawnie lub nie, odnosimy znaczenie i to my wypełniamy margines niezdeterminowania wynikający z niedoskonałości systemu rozpoznawczego rozpatrywanego urządzenia. Tego rodzaju antyindywidualizm interesuje Dennetta ze względu na to, że pokazuje on zależność intencjonalności wewnętrznej, a także pierwotnej od kontekstu. Wszelka wyróżniona w stanach wewnętrznych systemu lub organizmu intencjonalność posiada swe uzasadnienie i spełnienie swej prawdziwości w użyteczności dla dostosowania się organizmu do środowiska. Przyjmując **nastawienie intencjonalne** w celu zinterpretowania zachowania się organizmu musimy pamiętać o wyposażeniu jego stanów wewnętrznych w treści intencjonalne.

Mówiąc o intencjonalności stanów wewnętrznych umysłów ludzkich również musimy liczyć się z tym, co „Natura miała na myśli". Użycie takiego języka to antropomorfizacja, a ściślej mówiąc opis pewnego nieświadomego, mechanicznego procesu przy użyciu terminów intencjonalnych. Podkreślanym przez Dennetta znaczeniem teorii ewolucji jest usunięcie z niej umysłu, a znaczeniem teorii intencjonalności Brentano jest uznanie intencjonalności za oznakę tego, co mentalne. Dennett łączy te dwie teorie uzasadniając swój krok nieuniknionością użycia terminów mentalnych w języku nauk przyrodniczych. Podaje on przykłady prac z zakresu biologii, gdzie opisywane wirusy i enzymy zachowuje się intencjonalnie, a więc „wybierają", „popełniają błędy", „naprawiają", „unikają", „dążą". One same nie zachowują się tak, lecz my, chcąc je zrozumieć, przypisujemy im takie zachowania używając języka intencjonalnego.

---

[17] Tamże, s. 307.
[18] Tamże, s. 308-309.
[19] Tamże, s. 311-312.

Enzym naprawiający błąd w strukturze komórki nie rozpoznaje go jako błędu i może to oznaczać usunięcie intencjonalności z tego poziomu opisu. Niektórzy biologowie ewolucyjni nie przypisują wszystkim strukturom autonomicznym specyficznego znaczenia adaptacyjnego. Dzieje się tak, ponieważ takie struktury nie podlegają adaptacji do środowiska indywidualnie, a tylko jako część organizmu. „Treści" adaptacyjne można wtedy przypisywać do części organizmu pośrednio. Dopiero organizm jako całość zachowuje się zgodnie z intencją natury, a jego części, takie jak enzymy, są tu tylko narzędziami.

Złudzenie zawartej w procesie ewolucji inteligencji powstaje na skutek ograniczenia perspektywy postrzegania tego procesu. Wszelkie „nieinteligentne" posunięcia ewolucji, jako błędy nie służące przetrwaniu, zanikają. Ludzie są późnym, wysoce wyspecjalizowanym produktem naturalnej technologii, a wszelka intencjonalność umysłów ludzkich jest pochodną trwającego tysiące lat procesu. Jeżeli jest jakaś intencjonalność pierwotna, to jest to intencjonalność procesu doboru naturalnego.

1.5. Systemy intencjonalne

Dennett egzaminuje i wyjaśnia zachowanie się pewnych systemów przez przypisanie im przekonań, pragnień, nadziei, przeczuć i innych własności intencjonalnych. Dana rzecz jest **systemem intencjonalnym** [*intentional system*] tylko w relacji do podmiotu próbującego wyjaśnić i przewidzieć jej zachowanie. Do przyjęcia wobec tej rzeczy możliwe są trzy nastawienia.

Pierwsze z nich to **nastawienie projektowe** [*design stance*]. Jeżeli wiadomo jak system jest skonstruowany lub, używając metafory komputerowej, zaprogramowany, to można przewidzieć jego zachowanie się na podstawie stanów komputacyjnych. Przewidywanie takie oparte jest o pojęcie funkcji, jaką zachowanie się systemu spełnia. Jest więc ono celowo zależne, teleologiczne. Oznacza to, że projekt lub konstrukcja systemu składa się z większych i mniejszych części funkcjonalnych, a nastawienie projektowe zakłada, że każda z tych części funkcjonuje poprawnie. System taki przedstawić można w postaci schematu, w skład którego wchodzą połączone ze sobą elementy. Nastawienie projektowe stosować można na różnych poziomach opisu. Dla opisu funkcjonowania połączonych w rodzaj obwodu elementów wystarczy przedstawić ich funkcje i wzajemne relacje, bez potrzeby wnikania w wewnętrzną zasadę funkcjonowanie każdego z elementów z osobna. Elementy te można, na innym poziomie opisu, rozpatrywać przy użyciu tej samej zasady nastawienia projektowego.

Zwykle nastawienie projektowe przyjmuje się dla rozpatrywania „zachowania się" urządzeń mechanicznych. Dokonywane na tej zasadzie przewidywania polegają na znajomości samego tylko projektu konstrukcji, bez uwzględniania cech fizykalnych urządzenia. Nastawieniem, które uwzględnia aktualny stan fizykalny przedmiotu jest **nastawienie fizykalne** [*physical stance*]. W tym przypadku dla przewidywania zachowania się przedmiotu stosujemy znane nam prawa fizyczne. Nastawienie projektowe jest w stosunku do tego nastawienia pewnym uproszczeniem. Nastawienie fizykalne również mogłoby służyć do wyjaśnień działania komputera, lecz z uwagi na złożoność jego konstrukcji, zejście na poziom wyjaśnień fizykalnych uczyniłoby te wyjaśnienia zbyt skomplikowanymi.

W przypadku bardzo skomplikowanych programów komputerowych, do których należą programy szachowe, przewidywanie kolejnego ruchu komputera z poziomu projektowego

okazuje się niemożliwe. Lepiej jest potraktować komputer jako podmiot racjonalny, który pragnie wykonać najlepszy dla niego ruch. Racjonalność tego podmiotu nie oznacza nic więcej, niż jego optymalną dla osiągnięcia celu lub hierarchii celów konstrukcję. Przewidywanie zachowania się tego podmiotu uwzględnia ilość i rodzaj posiadanych przez system informacji. Tak jak przewidywanie oparte o nastawienie projektowe, takie przewidywanie zawodzi, gdy, nieuwzględniona w tym nastawieniu, struktura fizykalna ulega jakimś uszkodzeniom. Zawiedzie ono również, gdy ulegnie uszkodzeniom konstrukcja. Nastawienie to jest jednak najważniejsze dla koncepcji Dennetta. Zostało ono nazwane **nastawieniem intencjonalnym** [*intentional stance*]. Robione przy jego użyciu przewidywania są przewidywaniami intencjonalnymi i odnoszą się one do **systemów intencjonalnych**. Zakłada się, że systemy takie nakierowane są na osiągnięcie pewnych celów. Dennett stwierdza, że niewielką różnicę wprowadza nazwanie posiadanych przez te systemy informacji przekonaniami, a ich celów i podcelów pragnieniami. Uważa on tak, ponieważ pojęcie posiadania przez system informacji jest tak samo intencjonalne, jak pojęcie posiadania przez system przekonań.

Odróżnić należy posiadanie od zwykłego składowania informacji. Chodzi tu o posiadanie epistemiczne, zakładające wiedzę i poznanie informacji. Nie na miejscu są zarzuty poddające w wątpliwość to, że maszyna lub komputer mają rzeczywiście przekonania i pragnienia, gdyż Dennett niczego takie nie stwierdza. Zauważa on tylko, że przypisanie im cech intencjonalnych ułatwia czasem przewidywanie ich działania, a czasem jest ono jedynym możliwym postępowaniem. Decyzja, by przyjąć nastawienie intencjonalne, jest decyzją pragmatyczną. Można przyjmować nastawienie intencjonalne, jeżeli jest się przeciwnikiem komputera w grze w szachy, nastawienie projektowe, jeżeli jest się jego konstruktorem i nastawienie fizykalne, jeżeli poddaje się naprawie jeden z jego podzespołów.

Nastawienie intencjonalne związane jest z pewną intencjonalną teoria zachowania. Nasze zdroworozsądkowe wyjaśnienia zachowania się zwierząt są intencjonalne. Zakładamy ich racjonalność i założenie to jest tak silne, że w przypadku wystąpienia zachowań, z naszego punktu widzenia nieracjonalnych, przyczyny nieracjonalności poszukujemy w błędnym odczytaniu posiadanej informacji. Właściwe jej odczytanie powoduje, że interpretujemy zachowanie jako racjonalne. Zwierzęta, a nawet istoty pozaziemskie, powinny dzielić z nami przekonania, co do praw logiki. Przy opisie intencjonalnym, w którym takie przekonania im przypisujemy musimy przypisywać im również postępowanie zgodne z tymi prawami.

Możemy przypisywać im znajomość kilku reguł wynikania i przekonanie, co do wielu twierdzeń lub prawd, albo znajomość wielu reguł i przekonanie tylko co do ograniczonej ilości twierdzeń lub prawd logicznym. Tylko w idealnie logicznym systemie pojawią się wszystkie prawdy. W systemach, co do których stosujemy nastawienie intencjonalne nigdy nie występuje pełny ich zestaw, a ponadto nie muszą w nich być ważne wszystkie reguły wynikania. Dla coraz mniej doskonałych systemów odrzucamy coraz więcej prawd logicznych i możemy ostatecznie znaleźć się w sytuacji, kiedy lepsze efekty przyniesie przewidywanie z nastawienia projektowego. Dzieje się tak, kiedy odrzucamy założenie racjonalności systemu, co stosuje się w ekstremalnych wypadkach, takich jak opis zachowania się ludzi chorych psychicznie.

Za każdym razem, kiedy teoria proponuje nazwanie pewnego zdarzenia, stanu, czy struktury sygnałem, wiadomością, czy komendą nadaje mu ona pewną treść. Zakłada odczytanie sygnału i zrozumienie wiadomości lub komendy zgodnie z zakładaną w teorii inteligencją,

która nierozłączna jest z intencjonalnością. Behawioryzm Skinnera[20] zakładał, że próby wyjaśnień intencjonalnych będą bezwartościowe. Quine określał to mniej zdecydowanie obawiając się, czy próby takie znajdą kiedykolwiek zastosowanie. Metody proponowane przez behawioryzm nie eliminują jednak nastawienia intencjonalnego, a tylko je maskują. Próby wyjaśnienia zachowania się zwierząt w eksperymentach, przy użyciu nastawienia projektowego zawsze odwołują się do obecnych tam *implicite* intencji. Dennett przyznaje Skinnerowi rację odnośnie tego, że intencjonalność nie może być podstawą psychologii oraz, że właściwe jest poszukiwanie czysto mechanicznych regularności w aktywności badanego podmiotu. Nie należy jednak oczekiwać, że będą one bezpośrednio w jego zachowaniu widoczne.

Trzeba potraktować podmiot jako posiadający wejście i wyjście **system intencjonalny.** Należy go też wstępnie wyposażyć w pewną treść wyznaczającą mechanizm funkcjonowania systemu. Badania nad sztuczną inteligencją dostarczyły rozwiązań, w których wychodząc od intencjonalnego scharakteryzowania problemu przechodzi się do poszukiwań optymalnej dla zrealizowania intencji konstrukcji. Neurofizjologowie i psychofizycy również używają uwzględniającego treść nastawienia intencjonalnego i dopiero później przechodzą do nastawienia projektowego.

Koncepcja systemów intencjonalnych jest względnie jednolita. Nie zakłada świadomości, moralności, czy też boskiego pochodzenia rozpatrywanych systemów. Łatwiej jest rozpatrywać możliwość bycia maszyny systemem intencjonalnym, niż możliwość jej rzeczywistego myślenia, bycia świadomą lub moralnie odpowiedzialną. Jeżeli rozważamy jakiś podmiot jako posiadający umysł, duszę lub moralność, to podmiot taki zawsze będzie systemem intencjonalnym. Dennetta interesuje pytanie: „Co będzie prawdą o ludziach mających przekonania przez wzgląd na samo tylko ich bycie systemami intencjonalnymi, ze zdolnością do komunikowania się?”[21]

Systemy intencjonalne obdarzone językiem i zdolne do komunikowania się są subklasą systemów intencjonalnych. Dodatkowych narzędzi dla zdobycia wiedzy na temat umysłu i zdarzeń mentalnych może dostarczyć określenie ram teorii świadomości. Teorię taką Dennett wiąże ze zdolnością językową, którą rozpatruje w perspektywie teorii ewolucji, jako środek adaptacji do środowiska. Tylko komunikowanie przekonań prawdziwych ma wartość dla przetrwania jednostki lub gatunku. Zafałszowana komunikacja nie jest dobrym ostrzeżeniem innych systemów intencjonalnych o zbliżającym się niebezpieczeństwie i konieczności wybrania odpowiedniego w tym wypadku mechanizmu działania obronnego.

Prawdziwość przekonań definiowana jest w następujący sposób: „Jestem przekonany, że *p*.” implikuje „Powinno się być przekonanym, że *p*.” Jeżeli twierdzę, że *p* lub jestem przekonany, że *p* to zakładam, że mogę podać dowody prawdziwości *p* i dowody na to, że rzeczywiście jestem przekonany, że *p*. Moja wypowiedź powinna być więc adekwatna do stanu moich przekonań. Są to dwa warunki dotyczące prawdziwości przekonań i adekwatności do nich wypowiedzi. Na ogół warunki te uzupełniają się, ale mogą też być z sobą w konflikcie.

Jeżeli chcemy użyć koncepcji intencjonalności do rozpatrywania pewnego systemu musimy przyjąć nastawienie intencjonalne, zgodnie z którym system może mieć błędne przekonania. Dla systemów intencjonalnych, które mogą się komunikować, najlepszym sposobem sprawdzenia poprawności lub błędności ich przekonań jest ocena ich wypowiedzi. Systemy

[20] Por. D.C. Dennett, *Brainstorms*, s. 13-14.
[21] Tamże, s. 17.

takie nie są jednak perfekcyjnie racjonalne i mogą wypowiadać przekonania niezgodne z danymi empirycznymi, sprzeczne z innymi przekonaniami lub sprzeczne wewnętrznie. Należy więc przypisać im pewien zbiór przekonań nieracjonalnych.

Podejście fenomenologiczne zakłada, że introspekcja dostarcza danych niedostępnych badaniom zewnętrznym, prowadzonym z przyjęciem nastawienia intencjonalnego. Kiedy jednak podmiot dokonujący introspekcji komunikuje dane introspekcyjne nie wiemy, czy jego przekonania są słuszne, czy też błędne. Dla właściwej oceny musimy użyć danych innych niż fenomenologiczne i wyjaśnienie nasze przestaje być wyjaśnieniem fenomenologicznym. Wywołuje to problem oceny przekonań systemu. Pomocne w jego rozwiązaniu może być rozważenie pewnego zbioru transformacji **T** przekształcających jedne przekonania w inne. Należy zdeterminować **Ts** dla każdego systemu **S**. Wtedy wiedząc, że **S** jest przekonany, że *p* będziemy zdolni zdeterminować, o czym jeszcze przekonany jest **S**. Jeżeli system jest racjonalny idealnie, to przekonany będzie, co do każdej logicznej konsekwencji swych przekonań i nie będzie miał przekonań fałszywych. Wiemy, że żaden system intencjonalny nie jest racjonalny idealnie, więc zbiór transformacji **T**, dla każdego systemu wykazywał będzie pewne braki. Będzie on jednak zbiorem do pewnego stopnia spójnym i nie będzie zbiorem pustym.

Na gruncie takich właśnie filozoficznych przesłanek Dennett przechodzi do teorii ewolucji, która naturalistycznie przedstawia rozwój myślących i świadomych form życia. Zamierzonym celem jest dla niego pokazanie, jak problemy filozoficzne, które wciąż próbuje się rozwiązać przy użyciu niezrozumiałych, niedostępnych ludzkiemu poznaniu czynników, mogą zostać sprowadzone do postaci zagadnienia ostatniego stadium procesu doboru naturalnego, którym jest umysł ludzki. Zamiast wyjaśnień odwołujących się do cudu lub ingerencji sił nie poddających się analizie stosującej środki nauk szczegółowych, Dennett proponuje próbę zrozumienia ogromnego stopnia komplikacji teoretycznych i złożoności struktury fizykalnej związanych z mózgiem i umysłem. Mózg i umysł mogą być w koncepcji naturalistycznej utożsamiane. Nie znaczy to jednak, że Dennett redukuje zdarzenia mentalne do fizykalnych. Jego zdaniem świadczą one o występowaniu dwóch różnych rodzajów cech leżącego u ich podstawy systemu. Zrozumienie go wymaga wielu interdyscyplinarnych badań, które zaczynają się od prześledzenia procesu ewolucji na drodze doboru naturalnego.

2. Ewolucja

2.1. Wieża Generowania i Testowania

Dennett buduje Wieżę Generowania i Testowania[22], w której wyróżnia cztery poziomy. Pierwszy z nich zamieszkują istoty darwinowskie, drugi – istoty skinnerowskie, trzeci – istoty popperowskie i czwarty – istoty gregoriańskie. Nazwa ta wzięta została od nazwiska brytyjskiego psychologa, teoretyka roli informacji – Richarda Gregory'ego.

Zwierzęta, jako istoty „informacjożerne", wyczulone są na zawarte w materiale docierającym do ich zmysłów informacje. Na materiale takim pracuje wiele systemów i podsystemów, służących różnym rodzajom intencjonalności. Dla zobrazowania ewolucji zdolności

[22] Por. D.C. Dennett, *Natura Umysłów*, s. 99.

organizmów do znajdowania coraz lepszych strategii rozwiązywania problemów, Dennett proponuje schemat czterech poziomów Wieży Generowania i Testowania.

Najniższe piętro zamieszkują istoty darwinowskie, wykorzystujące do przetrwania sam tylko mechanizm ewolucji. Są one organizmami powstałymi w wyniku rekombinacji i mutacji genów. Powstałe na tym poziomie projekty, zgodnie z którymi zbudowane zostały organizmy, są dziełem przypadku. Istotna jest ich ilość i różnorodność. Tylko konstrukcje zbudowane na bazie najlepszych projektów genetycznych zdolne są przetrwać.

Jednym z występujących tu rozwiązań konstrukcyjnych są istoty zdolne przejawiać plastyczność fenotypową. Oznacza ona zdolność dostosowywania konstrukcji wewnętrznej do środowiska, w czasie rozwoju osobniczego. Odbywa się to dzięki wbudowanemu mechanizmowi wzmacniania posunięć właściwych. Istoty te Dennett nazwał istotami skinnerowskimi, ze względu na, podkreślane przez Skinnera, warunkowanie instrumentalne, uznawane za przedłużenie i uzupełnienie procesu doboru naturalnego. Istoty wzmacniające posunięcia przynoszące im skutki pozytywne podlegają procesowi uczenia się typu ABC[23]. Powtarzanie określonych zachowań wywołuje ponownie te same asocjacje, co skutkuje wytworzeniem odpowiednich połączeń neuronalnych. Nie różnią się one od połączeń przekazanych genetycznie i wraz z nimi tworzą strukturę wewnętrzną organizmu.

Nie wszystkie zdolności organizmu wyjaśnić można przy użyciu schematu ABC. Istnieje system doskonalszy, oparty na preselekcji zachowań i eliminacji zachowań błędnych. Istoty posługujące się tym systemem to istoty popperowskie, a zastosowane udoskonalenie polega na tym, by, jak to ujął K. Popper: „(...) nasze hipotezy umierały za nas."[24] Przeżycie istot skinnerowskich zagwarantowane jest szczęśliwym wykonaniem przez nie, posunięć właściwych jako pierwszych. Istoty popperowskie posługują się filtrem środowiska wewnętrznego, które zawiera wiele informacji o środowisku zewnętrznym i można w nim testować hipotezy, przed ich rzeczywistym użyciem. Samo myślenie o negatywnych posunięciach, pociąga za sobą myślenie o ich negatywnych skutkach. Zachodzi to bez wzmocnienia zewnętrznego i nazywane jest latentnym, czyli utajonym uczeniem się[25]. Tę zdolność wykorzystywania zaczerpniętych ze środowiska zewnętrznego informacji, do preselekcji swych możliwych zachowań, posiada wiele gatunków zwierząt.

Cechą wyróżniającą najwyższy poziom istot, nazywanych istotami gregoriańskimi, jest zdolność posługiwania się narzędziami. Narzędzia są nie tylko skutkiem działania inteligencji, lecz również istotnymi czynnikami jej rozwoju. Przyczyniają się one do niego, przekazując swoim użytkownikom część inteligencji potencjalnie zawartej w ich budowie i przeznaczeniu. Największe znaczenie ma potraktowanie słów jako narzędzi umysłu. Dzięki słowom oraz podobnym do nich narzędziom symbolicznym, istoty gregoriańskie konstruują coraz doskonalsze testery swych posunięć, wykorzystując do tego celu doświadczenie i inteligencję innych użytkowników języka. W ten sposób, na najwyższym poziomie Wieży Generowania i Testowania, dochodzi do wymiany i wspólnego użytkowania inteligencji całego gatunku.

[23] Z języka angielskiego: A – *associationism*, B – *behaviorism*, C – *connectionism*.
[24] D.C. Dennett, *Natura Umysłów*, s. 106.
[25] Tamże, s. 110.

2.2. Ewolucja w mózgu

Równoległe istnienie światów: nauk fizykalnych i nauk intencjonalnych jest, zdaniem Dennetta, nie do zaakceptowania. Nie można zredukować zdarzeń mentalnych, opisywanych językiem intencjonalnym, do zdarzeń fizykalnych, opisywanych językiem nauk fizykalnych. Opierając się na tradycyjnych założeniach intencjonalności, nie można podać jednej teorii, ujednolicającej świat i opisującej to, co mentalne i to, co fizykalne, przy użyciu tych samych środków. Jako jedyne wyjście uznaje Dennett, zmodyfikowanie tez intencjonalnych tak, by intencjonalność poddała się wnikliwszemu badaniu. Dennett proponuje osiągnięcie tego poprzez odrzucenie zewnętrznych zachowań i skupienie się na stanach i zdarzeniach wewnętrznych. Oświadcza, że jego teoria powinna odnieść stany i zdarzenia cerebralne, do treści i znaczenia.

Mówienie o sygnałach neuronalnych w mózgu, będących wynikiem pobudzenia organów zmysłów jest standardową praktyką neurofizjologów, lecz należy zastanowić się, co pozwala wyposażać te sygnały w określone treści i czy mogą to być treści porównywalne do tych, które zawarte są w naszych myślach. Ze względu na analogię opisu zachowania się umysłu ludzkiego i komputera, dla osiągnięcia swych celów Dennett proponuje użycie metafory komputerowej i zbadanie jej zastosowań.

Możliwość przejawiania zachowań intencjonalnych wymaga zdolności przechowywania informacji. Chcąc uniknąć mistycznych rozwiązań tego problemu, za przechowywanie informacji odpowiedzialną należy uczynić pewną strukturę materialną. Ponadto musi to być przechowywanie, nazywane przez Dennetta przechowywaniem inteligentnym [*ntelligent storage*] Przykładowo przechowywanie informacji przez komputer jest tylko zdolnością wydrukowania lub wyświetlenia na monitorze określonych liter lub symboli, w odpowiedzi na dane hasło. Sposób ten nie różni się specjalnie od staromodnego sposobu przechowywania informacji w książkach, na półkach bibliotek. Nazywany jest on przechowywaniem informacji tylko przez wzgląd na użytkowników komputera, którzy interpretują to, co pojawia się na jego wyjściu jako informacje. W podobnym sensie przechowują informacje pnie drzew, w postaci kręgów oznaczających lata lub pasma górskie, w postaci pokładów geologicznych oznaczających epoki. Nie przejawiają one zrozumienia zawartych w nich informacji, nie są przykładem przechowywania inteligentnego.

Przechowywanie inteligentne wyróżnia się tym, że system, w którym ono zachodzi zdolny jest zrobić z informacją coś więcej, niż samo tylko okazywanie formy jej zapisu. W wielu wypadkach trudno jest powiedzieć, co należałoby uznać za *użycie* informacji, ale pewne programy komputerowe robią, zdaniem Dennetta, z informacją dostatecznie wiele, by uznać je za przykłady przechowywania inteligentnego. Przechowywanie takie bazuje na przechowywaniu nieinteligentnym i jest przechowywaniem „dla systemu”, nie tylko dla jego użytkowników. System musi więc mieć pewne potrzeby, dla zaspokojenia których używa tej informacji. Dennett stwierdza, że kryterium przechowywania inteligentnego może być tylko ocena stosowności zachowań systemu, w odpowiedzi na pewne zewnętrzne warunki początkowe, z których system czerpie informację na wejściu. Dostosowanie takie nie jest wyrażane fizykalną ani formalną wewnętrzną [*ntrinsic*]charakterystyką rzeczy ani zdarzenia. Charakterystyka wewnętrznych zależności zawartych pomiędzy wejściem, a wyjściem systemu może okazać się niewystarczającą dla wyjaśnienia, czym jest inteligencja. Dennett zauważa, że przechowywanie inteligentne nie daje się wyjaśnić na podstawie samych tylko procesów mechanicznych przechowywania informacji, charakterystycznych dla komputerów i

im podobnych nośników informacji. Jest ono dodatkową i oddzielną zdolnością, dla której stara się on pokazać warunki możliwości zaistnienia.

Praca mózgu powinna wywoływać zachowania dostosowane do środowiska. Jeżeli odrzucić możliwość całkowicie przypadkowego dostosowywania się organizmu, to mózg musi być wyposażony w zdolność rozpoznawania znaczenia pojawiającej się na wejściu informacji. Jednak żadne fizykalne zdarzenia w mózgu nie są wyposażone w znaczenia wewnętrzne. Włókna nerwowe ani komórki nie przesądzają o znaczeniu wiadomości, jaką niosą z sobą przebiegające przez nie sygnały elektryczne. Dennett wyciąga stąd wniosek, że wyjaśnianie posiadania przez nie znaczeń nie może zachodzić dzięki samym tylko strukturom elektro-chemicznym. Mózg jest ślepy na, dostarczające informacji na jego wejście, warunki zewnętrzne. Dennett stwierdza, że skoro rozpoznawanie znaczenia nie jest oparte o jakikolwiek test fizykalny, to mózg musi jednak to robić na drodze przypadku. Ścieżka połączeń neuronalnych prowadzących od zakończeń aferentnych (wejście) do zakończeń eferentnych (wyjście) powstaje dzięki przypadkowym próbom i eliminowaniu połączeń błędnych. Jeżeli odpowiedzią na zdarzenie lub bodziec wejściowy jest właściwe zachowanie się organizmu, ścieżka połączeń neuronalnych zostaje rozpoznana i zachowana przez mózg w postaci odpowiednich struktur funkcjonalnych.

Strukturą funkcjonalną może być jakakolwiek cząstka materialna. Zasada wyboru, zachowania i używania struktur funkcjonalnych przydatnych dla przetrwania organizmu, a odrzucania lub powstrzymywania rozwoju struktur szkodliwych nie może, zdaniem Dennetta, sprowadzać się do jakiejś ich cechy fizykalnej. Nie jest tak, że struktury szkodliwe giną, gdyż cierpią na fizykalny defekt, a zachowane i utrwalone zostają struktury korzystne, gdyż posiadają sprzyjającą temu cechę fizykalną. Dla wyjaśnienia działania zasady wyboru właściwych struktur funkcjonalnych Dennett proponuje zastosowanie hipotezy ewolucyjnej. Dla jej zobrazowania podaje on prosty przykład trzech organizmów: A B C, o nieskomplikowanych systemach nerwowych, różnie reagujących na pewien określony bodziec. Bodziec ten zmusza organizm A do odwrotu, powoduje niezdecydowanie organizmu B, a organizm C skłania do zbliżenia się w kierunku źródła sygnału. Może nim być zapach pożywienia lub wróg. W zależności od tego oraz od reakcji zdeterminowanej przez system nerwowy, każdy z organizmów ma inną szansę na przetrwanie. W ten sposób prosty system nerwowy, mający tylko trzy możliwe stany i niezdolny do rozpoznania znaczenia sygnału gwarantuje przetrwanie organizmów najlepiej dostosowanych do środowiska. Niepotrzebna jest znajomość znaczenia, gdyż testowane są wszystkie możliwe stany na wyjściu i automatycznie zachowane zostają te z nich, które sprawdzają się najlepiej.

W procesie ewolucji największą szansę na przetrwanie mają te organizmy, które reagują różnie na różne bodźce, a więc mają reakcję ku bodźcom pozytywnym, reakcję odwrotu na bodźce negatywne i reakcje niezdecydowania na bodźce niezinterpretowane. Reakcje te utrwalane są w rozwoju gatunkowym i przekazywane kolejnym pokoleniom genetycznie. Przykładem jest tutaj połączenie aferentno-eferentne gwarantujące reakcję odwrotu na wejściowy sygnał bólu. Dennett pisze, że teoretycznie możliwy jest wytwór ewolucji, który ma wrodzone, zagwarantowane genetycznie wszystkie potrzebne mu do przetrwania połączenia aferentno-eferentne. Organizm taki jest niezdolny do uczenia się nowych zachowań, ale on takich zachowań nie potrzebuje. Wytworzenie przez ewolucję takiego organizmu mogłoby jednak nastąpić tylko w niezmiennych warunkach środowiskowych, w których przetrwanie gwarantuje stały zestaw reakcji. Przeniesienie takiego organizmu do innego środowiska byłoby równoznaczne ze skazaniem go na wymarcie. Następowanie

trudnych do przewidzenia zmian jest jedną z cech środowiska naturalnego, więc organizmy z ostatecznie już skompletowanym zestawem struktur funkcjonalnych nie mogą przetrwać.

Zbyt wielka ilość wrodzonych struktur funkcjonalnych okazuje się dla organizmu szkodliwa, lecz pewna ich ilość, pisze Dennett, jest konieczna. Sygnał aferentny, sam przez się nie mówi, że zadziała jako pozytywne lub negatywne sprzężenie zwrotne [*eedback*]. Reakcje na pewne sygnały aferentne muszą być wrodzone, a problemem jest pokazanie, w jaki sposób te wstępnie już zawarte w strukturze wewnętrznej, schematy zachowań pozwalają mózgowi uczącego się organizmu na dokonywanie stosownych rozróżnień wśród nowych dla organizmu pobudzeń.

Dla wyjaśnienia tego problemu konieczne jest, zdaniem Dennetta, odrzucenie tradycyjnego obrazu mózgu, jako organu genetycznie wyposażonego w dwie strony: aferentną (wejście) i eferentną (wyjście), z pewną ilością wrodzonych połączeń pomiędzy tymi dwoma stronami (odruchy, instynkty, tropizmy) oraz przestrzenią połączeń możliwych, z których odpowiednie zostają wybrane przy udziale aferentnej analizy i skoordynowania eferentnego. Ten sposób wymaga „małego człowieka w mózgu", który rozumie i dokonuje właściwych wyborów, a więc inteligentnie używa mózgu. Przenosiłoby to wyjaśnienia na płaszczyznę samego tylko umysłu, a tego Dennett chce uniknąć. Zastosowanie homunculusa nie wyjaśnia działania mózgu, który jest przedmiotem jego zainteresowań.

Dennett nawiązuje do koncepcji zachowań operacyjnych [*operant behavior*][26] B.F. Skinnerra. Zachowania te nie są wybierane, lecz po prostu emitowane przez eferentną stronę mózgu. Są to więc przypadkowe zachowania, które w jakiś sposób zostają następnie połączone z właściwymi bodźcami. Dennett proponuje nie poszukiwać sposobu połączenia bodźca z właściwym zachowaniem, lecz założyć wstępne pobudzanie wszystkich możliwych zachowań operacyjnych. Problemem jest wtedy usunięcie połączeń nieprzydatnych. Potrzebna do tego jest, zdaniem Dennetta, wewnątrzcerebralna funkcja spełniająca taką samą funkcję, jaką spełnia ewolucja w procesie rozwoju gatunkowego, w którym nieprzystosowane gatunki wymierają. Dla zobrazowania tego procesu potrzeba pewnych informacji na temat budowy mózgu.

Liczba neuronów w ludzkim mózgu jest rzędu 10,000,000,000. Mają one wielokrotne wejścia – dendryty i pojedyncze wyjście – akson, który może być połączony z małą lub wielka liczbą innych komórek nerwowych. Pomiędzy zakończeniami wyjściowymi odgałęzień aksonu, a wejściowymi dendrytami lub również biorącym udział w połączeniu ciałem komórki – soma, istnieje przerwa nazywana synapsą. Impulsy przewodzone przez włókno nerwowe przekraczają ją tylko wtedy, jeżeli osiągną one pewną minimalna częstotliwość. Dla koncepcji Dennetta istotny jest fenomen wartości progowej [*threshold*]. Impulsy wejściowe mogą przyspieszać lub powstrzymywać wyjściowe wyładowanie neuronu. Następuje ono, jeżeli suma tych impulsów przekroczy określoną dla neuronu wartość progową. Wartość ta może być różna dla różnych neuronów i jest obniżana częstym wyładowywaniem neuronu, a podwyższana jego aktywowaniem rzadkim. Mechanizm ten wyjaśnia utrwalanie w mózgu właściwych połączeń.

Nie wyjaśnia to jednak przypisania sygnałom neuronalnym treści lub znaczenia. Jeżeli pewien neuron przewodziłby sygnał dalej tylko w przypadku pobudzenia sygnałem wywołanym określonym wzorem zapisanym na siatkówce oka i doprowadzonym do niego nerwem

[26] Por. B.F. Skinner, *Science and Human Behavior*, 1965, s. 59-90.

optycznym, to byłoby to działanie jednoznaczne. Jednak poza peryferyjnymi neuronami zakończeń aferentnych, neurony działają wieloznacznie, ich wyładowania może powodować wiele różnych sygnałów, co jest cechą zapewniającą właściwe funkcjonowanie mózgu. Błędne funkcjonowanie neuronu bliskiego zakończeniom aferentnym powoduje tylko niewielkie ograniczenie dostępu do informacji wejściowej, lecz usterka neuronu w dalszej części łańcucha przekazywania informacji zsumowanej z wszystkich neuronów poprzedzających, mogłaby powodować znaczne zaburzenie funkcji mózgu. Dlatego mózg duplikuje ścieżki połączeń, co gwarantuje niezawodność. Jest ona większa niż niezawodność jego części składowych. Dla zobrazowania konieczności dublowania połączeń Dennett przytacza kalkulację, którą podał Arbib[27]. Rozważany jest w niej łańcuch o *n* modułach (neuronach) oraz prawdopodobieństwo *p* usterki pojedynczego neuronu. Prawdopodobieństwo pojawienia się poprawnej informacji na wyjściu całego łańcucha wynosi $(1 - p)^n$. Dla p = 1% i n = 70 wynosi ono 0,5. Wynik ten będzie o wiele mniejszy, jeżeli weźmie się pod uwagę ilość neuronów w mózgu. Dennett opowiada się za większym prawdopodobieństwem prawidłowego funkcjonowania neuronu, dublowaniem łańcuchów przenoszących tę samą informację i mechanizmami kontroli poprawności transmisji na różnych poziomach przekazu, co da w efekcie wysokie prawdopodobieństwo uzyskania poprawnej informacji na wyjściu.

Postawiony przez Dennetta problem dotyczy wybierania przez mózg stosownych połączeń aferentno-eferentnych spośród wielu połączeń przypadkowych. Rozwiązując go należy wziąć pod uwagę złożoną naturę sygnałów neuronalnych oraz zasadę działania ewolucji, w której wystąpić musi konflikt pomiędzy pewnymi cechami środowiska a gatunkami, które mają zostać wyeliminowane. Odpowiednikiem tej zasady w mózgu jest niezgodność pewnych struktur funkcjonalnych ze schematem połączeń wrodzonych, które muszą pozostać niezmienione. Są one związane z zachowaniami użytecznymi i korzystnymi dla gatunku w całym jego procesie ewolucyjnym. Muszą one również mieć zdolność powstrzymywania rozwoju i usuwania struktur z nimi niezgodnych. Ma to odbywać się tak, jak w ewolucji gatunków, nie przez wyginięcie wszystkich przedstawicieli (w mózgu odpowiednich łańcuchów neuronalnych), lecz przez ich niezdolność do reprodukcji. Wewnątrzcerebralny proces ewolucyjny uniemożliwia powielanie struktur wcześniej odrzuconych.

Całkowita redukcja wpływu impulsów pochodzących ze środowiska jest niemożliwa, gdyż są one potrzebne choćby po to, by zapoczątkować działanie systemu. Z drugiej strony mózg jako organ fizyczny nie mógłby wytwarzać żadnych nowych struktur wewnątrzcerebralnych, gdyby nie posiadał pewnych struktur początkowych umożliwiających dokonywanie wyborów i tworzenie konstrukcji. Struktury te nie muszą objawiać się w konkretnych zachowaniach. Równie skuteczna dla przetrwania jest ich obecność potencjalna. Zaczynają one funkcjonować dopiero w określonych sytuacjach, w których łączą właściwą reakcję organizmu z przyjemnością, a niewłaściwą z bólem. Ból i przyjemność spełniają funkcje kary i nagrody w procesie uczenia się organizmu, chociaż przy działaniach złożonych nie muszą one być bezpośrednio związane z jego kolejnymi posunięciami.

Wyjaśniając proces uczenia się i rozróżniania w mózgu, zakłada się powstawanie struktur, które są tam z pewnego powodu. Dennett zwraca uwagę na istotność rozróżnienia na istnienie powodu do posiadania przez organizm struktury i na posiadanie przez organizm powodu do jej posiadania. Nie można powiedzieć o człowieku, że ma powód do posiadania określonej struktury neuronalnej. Istnieją określone powody jej wystąpienia, ale nie można powiedzieć,

[27] Por. D.C. Dennett, *Content and Consciousness*, s. 55.

że to człowiek ma te powody. Powodem istnienia struktury neuronalnej są określone warunki użyteczne dla przetrwania organizmu. Przy zaistnieniu warunków odmiennych dana struktura nie spełniłaby swej roli i musiałaby zostać zastąpiona inną, gwarantującą lepsze dostosowanie organizmu do środowiska. Jest to teleologiczne wyjaśnianie warunków istnienia struktur neuronalnych, a Dennett chce pokazać, że można ten sposób wyjaśniania zamienić na nieteleologiczny, używający kolejnych posunięć procesu ewolucyjnego.

Krytycy teorii zachowania typu bodziec – reakcja często przeciwstawiali takie zachowanie zachowaniu celowemu. Dennett zadaje pytanie: „Czy struktury posiadające powód swego istnienia mogą zostać użyte do wyjaśnienia zachowań nakierowanych na cel [*goal directed*]?” Chcąc ustanowić kryteria opisu zachowań celowych wychodzi on od stwierdzenia, że słowo „cel” zastosowane do opisu działania programów komputerowych, jest przesadne. Jeżeli cel ten osiągnąć można w skończonej ilości kroków, to posiada go programista, a komputer jest tylko narzędziem. Jeżeli jednak zadany problem jest złożony i cel może zostać osiągnięty tylko na drodze wielu różnorodnych posunięć, to zdaniem Dennetta wygląda to na sytuację, w której uznanie, że komputer jest nakierowany na dany cel wydaje się bardziej stosowne. Podany zostaje tu przykład programu komputerowego GPS[28] [*general problem solver*], który składa się z pewnej ilości podprogramów dokonujących przekształceń matematycznych sygnałów wejściowych oraz z systemu zapewniającego przejścia z jednego podprogramu do drugiego. Zadanie GPS problemu do rozwiązania wymaga podania wyrażenia wejściowego oraz określenia wymaganej formuły wyjściowej. Zwykle na wejście podaje się zbiór zdań logicznych, a na wyjściu oczekuje się ostatniej linijki przeprowadzonego dowodu. Podanie jej jest wtedy celem komputera. Osiągnięcie go odbywa się na drodze wielu kroków próbnych i pośrednich. Działanie takie nazywa Dennett tylko określonym przez cel [*goal terminated*].

Nakierowanie na cel wymaga dalszych wyjaśnień. Stwierdzamy je u zwierząt na podstawie zaobserwowanych wzorów zachowań, które równie dobrze mogą być tylko określone przez cel. Różnicą pomiędzy tymi dwoma proponowanymi rodzajami opisów zachowań ma być znajomość celu – w przypadku zachowań nakierowanych na cel i jego nieznajomość – w przypadku zachowań celem określonych.

Problemem jest również określenie znajomości lub nieznajomości celu. Zwykle wiedza o tym, że *p* wymaga stwierdzenia, że *p* i pociąga za sobą możliwość zakomunikowania, że *p*. Zachowania celowe zarezerwowane byłyby więc tylko dla użytkowników języka. Nie przekonuje to jednak Dennetta, który zachowania takie przypisuje również małym dzieciom i niektórym zwierzętom. Musi on więc poszukiwać innego kryterium wiedzy i proponuje zastąpienie nakierowania na cel określeniem przez cel, który nie musi być z góry znany. Obserwacja zwierząt nie upoważnia, zdaniem Dennetta, do przypisywania im innego rodzaju zachowań jak tylko zachowań określonych przez cel, charakterystycznych też dla systemu GPS. Autor zaznacza, że sprowadzenie zachowań ludzkich do tego poziomu stanowi odrębny problem, wymagający rozpatrzenia w perspektywie ewolucji gatunku ludzkiego. Należy tu wziąć pod uwagę wymagany stan na wejściu systemu, którym mogą być peryferyjne wrażenia odbierane przez organy zmysłowe. Jednak u ludzi i zwierząt rzadko celem jest coś takiego, jak określony wzór na siatkówce oka. Cel ten jest na ogół bardziej złożony niż dane pobudzenie sensoryczne i powinien być raczej opisywany bardziej ogólnie, jako osiągnięcie określonych warunków zewnętrznych. Dennett wspomina, że doświadczenia z urządzeniami służącymi do rozpoznawania obrazów, zbudowanymi z modułów podobnych do neuronów, lecz o wiele prostszymi w budowie niż sieć neuronalna mózgu człowieka, wykazały

[28] Tamże, s. 65.

możliwość dokonywania przez takie sieci uogólnień. Jest to równoznaczne z możliwością określenia stanu wyjściowego systemu jako uogólnienia warunków zewnętrznych.

Przykładowo, kiedy na siatkówce oka pojawia się wzór linii pionowych, wywołuje on związany z nim peryferyjny sygnał neuronalny. Na wyższym poziomie sieci neuronalnej następuje wyładowanie neuronów oznaczające otoczenie zwierzęcia przez przedmioty pionowe. Na piętrze kolejnym pojawia się sygnał oznaczający, że zwierzę ukryło się pomiędzy drzewami. Taka aktywność wyższych poziomów sieci neuronalnych powinna być, zdaniem Dennetta, uznana za stan końcowy, który jest celem zachowania nakierowanego celowo. Rozwijana przez niego centralistyczna teoria zachowań wyznacza kierunek badań, lecz nie podaje ostatecznego rozwiązania problemu przypisywania treści lub znaczenia stanom mózgu.

2.3. Prawo efektu

W *Science and Human Behavior* Skinner utożsamił zaproponowane przez siebie zachowania operacyjne z wprowadzonym przez Thorndike'a prawem efektu [*The Law of Effect*][29]. Prawo to mówi o wpisaniu lub, jak to się często nazywa, wdrukowaniu [*inprinting*] wzorów pewnych zachowań wyznaczających krzywą uczenia się. Do ich zachowania i powtarzania wyznaczonych nimi zachowań konieczne jest wzmocnienie w postaci nagradzania właściwych posunięć. Dennett podkreśla podobieństwo tego prawa do mechanizmu działania ewolucji. Pisze on o istnieniu par: <bodziec, reakcja>, których przetrwanie zagwarantowane jest ciągłym wzmacnianiem nagrodą. Występujące obok nich inne pary nie zostają reprodukowane, gdyż nie podlegają wzmocnieniu.

Ewolucja wyjaśnia świat mechanistycznie, ale ponadto teleologicznie, całkowicie zależnie od celu i znaczenia, którym jest przetrwanie. Dennett zwraca uwagę, że decyzja na przyjęcie tego rodzaju materializmu nie wynika ze skłonności, czy nastawień, lecz z chęci uniknięcia wielu problemów i pytań, które pojawiają się, kiedy przyjęte zostaje inne stanowisko. Materializm teorii ewolucji Darwina nie musi być jedyną możliwą teorią, nie wywołującą nadmiernej ilości pytań i problemów na temat świata istot żywych. Jest on jednak jedyną taką teorią dotychczas poznaną i to uważa Dennett za wystarczający powód, by tę teorię przyjąć. Uznaje on jednak niewystarczalność zastosowania prawa efektu dla całkowitego wyjaśnienia procesu uczenia się. Pisze o przekonaniu, że prawo efektu nie jest po prostu dobrym rozwiązaniem, ale, że jest w ogóle rozwiązaniem jedynym możliwym. Wierność temu prawu, w jego behawiorystycznej lub peryferyjnej wersji, skłoniła psychologów do omijania wielu problemów, aby ocalić ich teorię. W przypadku przejawiania przez organizm całkowicie nowego zachowania inteligentnego postulowali oni istnienie odpowiedniego, wzmacniającego takie zachowanie doświadczenia w indywidualnej historii tego organizmu. Jest to, zdaniem Dennetta, tylko naśladowanie hipotezy rozwoju gatunkowego, przeniesione na rozwój jednostkowy. W związku z tym proponuje on odmienną od behawiorystycznej wersję prawa efektu.

Zgodnie z tym, co pisze Dennett istotne jest, że prawo efektu i zasada doboru naturalnego nie tylko są analogiami, ale zostały zaprojektowane po to, by działać wspólnie. Dostosowanie do środowiska, przejawiające się w wytwarzaniu takim jak budowanie gniazda czy pajęczyny,

[29] Por. B. F. Skinner, *Science and Human Behavior*, s. 60.

należy wyjaśniać przy użyciu tej samej zasady, co dostosowanie przez posiadanie odpowiednich narządów, takich jak skrzydła czy oczy. Istoty wyposażone w tropizmy i instynkty przydatne w ich niszy środowiskowej, mają większą szansę przetrwania, niż istoty nimi nie obdarzone. One też zostają wybrane na drodze doboru naturalnego. Zachowania oparte na tropizmach nie są plastyczne. Organizmy całkowicie podlegające tropizmom nie mogą więc być ostateczną odpowiedzią ewolucji na zmienne warunki środowiskowe. Sprawdzać się w nich będą najlepiej te istoty, które będą posiadały zdolność odróżniania korzystnych modyfikacji zachowań od niekorzystnych i utrwalania tych pierwszych.

Dennett proponuje wprowadzenie pewnej klasy zdarzeń wewnętrznych organizmu, uwarunkowanych genetycznie, które mają za zadanie zwiększać prawdopodobieństwo wystąpienia zdarzeń spełniających rolę kontroli zachowań. Zdarzenia te nazywa zdarzeniami wzmacniającymi. Szansę na przetrwanie mają tylko te mutacje, które wyposażone zostają w odpowiednie zdarzenia wzmacniające, rozumiane jako zdolność uczenia się. Uczenie się jest tutaj zmianą następującą w środowiskowo uwarunkowanym kierunku. Do grupy zachowań wrodzonych (instynkty, tropizmy) dodane zostają więc pewne zachowania wzmacniające. Skinner pisał o tym, że gdzie nie wystarczają zachowania dziedziczone genetycznie, użyta zostaje dziedziczona genetycznie zdolność do modyfikowania zachowań. Dlatego Dennett nazywa istoty posiadające tę zdolność istotami skinnerowskimi. Zgodnie z prawem efektu istoty te podlegają bezpośredniemu nagradzaniu ze środowiska zewnętrznego, które działa jak sprzężenie zwrotne o działaniu wzmacniającym.

Wzmacnianie to ma inną naturę w przypadku istot popperowskich. Dla nich jest ono gwarantowane sprzężeniem zwrotnym ze środowiskiem wewnętrznym, co zapewnia bezpieczną wersję uczenia się. Behawioryści, pisze Dennett, mają podejście peryferyjne i nie rozpatrują zastosowania prawa efektu poprzez środowisko wewnętrzne, co on proponuje zrobić. Określa środowisko wewnętrzne, jako ewoluujące w jednostkach w efekcie warunkowania operacyjnego. Opisuje je przy użyciu nieintencjonalnego języka. Jest to: „(...) jakakolwiek strefa wewnętrzna, która może oddziaływać i podlegać oddziaływaniom cech systemów kontroli potencjalnych zachowań”[30]. Wybrane zostają w tym środowisku takie jego elementy, które zawierają potencjalne relacje przyczynowo – skutkowe mające największe prawdopodobieństwo użyteczności dla dostosowania się organizmu do środowiska. Dennett zastanawia się, jakie warunki muszą spełniać fragmenty mózgu, by zapewniać lepszy, niż tylko przypadkowy, mechanizm ich wyboru. Pisze on, że polegamy tu na cechach fizykalnych mózgu, które opisywane są językiem fizykalnym i podlegają pewnym mechanicznym zasadom. Systemy kontroli potencjalnych zachowań organizmu natomiast, spełniają rolę funkcjonalną. Dla wybranych typów zdarzeń fizykalnych w mózgu i ich roli funkcjonalnej powinna zostać określona normalna [*normal*] lub systematyczna [*systematic*] wzajemna relacja. Scharakteryzowana fizykalnie sieć połączeń neuronalnych składać się będzie z egzemplarzy scharakteryzowanych fizykalnie i spełniać będzie również określone funkcje. Tak jak w warunkowaniu operacyjnym, zostaną one wzmocnione, gdy wywarte zostanie pożądane oddziaływanie na środowisko.

Dennett podkreśla trudność występującą w omawianiu tych funkcji przy użyciu języka behawioralnego i, ze względu na odgrywanie przez egzemplarze fizykalne określonej roli w organizacji fizykalnej, proponuje przypisanie im znaczenia lub treści. Jako przykłady elementów kontrolujących potencjalne zachowania podaje on mapy, prawdziwe przekonania, słuszne oczekiwania i dobre pomysły, a więc narzędzia psychologii poznawczej. Chce

[30] D.C. Dennett, *Brainstorms*, s. 79.

pokazać, że psychologia poznawcza jest ściśle związana z prawem efektu i w tym celu przytacza przykłady z badań nad sztuczną inteligencją. Postępowanie przyjęte przez behawiorystów prowadziło od pewnych zaobserwowanych zasad mechanicznych do opisów intencjonalnych. Badacze sztucznej inteligencji wychodzą od problemów intencjonalnych, np. „Jak sprawić, by komputer rozumiał pytania?”, rozbijają je na podproblemy, a następnie sprowadzają je na poziom struktur czysto mechanicznych. Wyróżnione podsystemy zajmują się scharakteryzowanymi intencjonalnie zadaniami, takimi jak: pamiętanie, rozróżnianie, porównywanie. Działania te opisuje Dennett przy pomocy metafory homunculusów posiadających informacje, cele i strategie. Każdy z nich rozbity zostaje na mniejsze i mniej mądre homunculusy, aż do homunculusów potrafiących tylko dodawać lub odejmować, a te mogą być zastąpione przez maszyny. Potwierdzeniem tego przejścia od antropomorfizacji do wyjaśnień czysto mechanicznych jest każde uruchomienie programu komputerowego, przy pisaniu którego programista stosował, na pewnym poziomie, nastawienie intencjonalne.

Znaną strategią w programowaniu jest używanie podprogramów generujących i testujących, co przypomina sposób działania doboru naturalnego. Prawo efektu nagradzające właściwe zachowania organizmów w środowisku zewnętrznym, tutaj przejawia się, jak to rozumie Dennett, jako nagradzanie właściwych rozwiązań dobrym działaniem programu. Taka wersja prawa efektu pozwala, zdaniem Dennetta, na wyciągnięcie większej ilości wniosków, niż wersja behawiorystyczna. Badacze sztucznej inteligencji dokonują eksperymentów, w których zajmują się symulacją zdolności poznawczych żywych organizmów. Konstruują oni również systemy nie wzorowane na żadnym rodzaju inteligencji naturalnie występującej w naturze. Sztuczną inteligencją nazywane są więc wszelkie możliwe, realizowalne mechanicznie typy inteligencji. Nie ma znaczenia, w jakiej postaci zrealizowany zostanie mechanizm. Ważne jest by funkcjonował on zgodnie z projektem, a każdy jasno przedstawiony projekt może zostać zrealizowany. Jasne przedstawianie spełniane jest w odpowiednim języku programowania, a więc można je przedstawić jako stany maszyny Turinga. Programy takie, pisze Dennett[31], stanowią analogię do występujących w naturze tropizmów. Problemem jest uczynienie takich programów samoprogramujacymi się [*self-design*], co byłoby analogią do naturalnego procesu uczenia się. Dennett twierdzi, że można to zrealizować tylko na zasadzie generowania i testowania, a więc w pewnej formie nagradzania. Prawo efektu ma swoją analogię w teorii sztucznej inteligencji, teorii informacji czy teorii samoorganizujących się systemów. Proces uczenia się Dennett określa jako proces samoprojektowania się i wyjaśnia, w jaki sposób powstają nowe projekty.

Odróżnienie nowopowstałych projektów od innych wymaga użycia przez system informacji. Informacja ta pochodzić może z zewnątrz systemu, z jego wnętrza lub z obu tych przestrzeni. Jeżeli pochodzi ona tylko z wnętrza systemu, to musi być już zawarta w jego poprzednich projektach i nie zachodzi wtedy uczenie się. Istnieje jedynie system lub organizm z gotowymi już projektami. W obu pozostałych przypadkach zachodzi rzeczywiste uczenie się, wykorzystujące zupełnie nowe dla systemu lub organizmu informacje. Nowopowstające projekty nie są wtedy zdeterminowane przez same tylko projekty znane wcześniej. Te wcześniejsze projekty muszą jednak zapewnić zdolność wybierania odpowiednich nowych projektów spośród wielu projektów próbnie generowanych. Jeżeli stany wewnętrzne tego nie gwarantują, to mamy do czynienia z automatem niezdeterminowanym[32]. W przypadku takiego automatu przeprowadzenie przez niego kolejnych operacji musi zaczekać, aż wygenerowane i przetestowane zostaną pewne rozwiązania próbne. Zdeterminowanie automatu następuje poprzez podanie dodatkowej informacji na jego wejście.

[31] Tamże, s. 84.
[32] Tamże, s. 85.

Analogie pomiędzy automatami i procesem doboru naturalnego pokazują, że w opisie działania systemów intencjonalnych nie da się uniknąć prawa efektu. Ważna jest tutaj wysoka selektywność testerów, a dokonywanie przez nie właściwych wyborów zależne jest od wyposażenia genetycznego systemu lub od jego wyposażenia genetycznego wraz z historią rozwoju indywidualnego. Dennett zwraca uwagę, że trudno jest określić, która z tych możliwości rzeczywiście zachodzi. Bez względu na to, która nią będzie, w jej działaniu ważniejsze od przypadkowego generowania rozwiązań jest wybieranie rozwiązań właściwych. Samo testowanie przebiega zwykle w umyśle ludzkim podświadomie. Świadomości ujawniony zostaje tylko jego wynik. Zdarzają się jednak stany świadomości, w których podmiot może mieć pewien wgląd w kształtowanie się konkretnej myśli. Dennett podaje ten przykład dla uzasadnienia swych poglądów. Chodzi tu o stany, kiedy pojawia się w umyśle wiele myśli zanim wybrana zostanie ta właściwa. Jest to rodzaj testowania, który Dennett proponuje dla wyjaśnienia działania inteligencji.

2.4. Ewolucja świadomości

Dennett rozpoczyna poszukiwania źródeł świadomości od stwierdzenia, że świadomość nie zawsze istniała w postaci takiej, w jakiej objawia się ona w umysłach ludzkich. Musiała się więc ona rozwinąć z elementów, które nią nie były, a prześledzenie przekształceń prowadzących do wytworzenia się jej fenomenu w pełni, powinno pomóc w zrozumieniu jej natury. Należy odrzucić „czarna skrzynkę" świadomości obecną w behawioralnym badaniu zachowań zewnętrznych i zająć się świadomością od wewnątrz. Zamiast prób dociekań mechanizmów w mózgu, które mogłyby wyjaśnić dane zachowanie Dennett proponuje rozważenie ewolucji mechanizmów mózgowych, od prostych do coraz bardziej złożonych i stwierdzenie, które z nich będą mogły wyjaśnić skomplikowane schematy zachowań. Rozpoczyna od rozróżnienia na powody istnienia świadomości, którym przypisuje treść, a więc zakłada istnienie umysłu oraz przyczyny traktowane czysto mechanicznie, w żadnym sensie nie mentalne.

Świadomość zaczęła rozwijać się w świecie, w którym były tylko przyczyny, nie było teleologii ani żadnego rodzaju dóbr. Od momentu, w którym pojawiły się pierwsze proste samoreproduktory, mówić można o celu, którym była samoreprodukcja. Dobrem było wszystko, co jej służyło. Samoreproduktory, by zapewnić sobie przetrwanie, musiały nauczyć się rozpoznawać to, co przetrwaniu służy, co mu zagraża i co jest mu obojętne. Kiedy zaczęły one chronić swe istnienie musiały wyznaczyć linię podziału pomiędzy sobą a światem. Musiały więc, pisze Dennett, stać się egoistyczne. Ich systemy rozpoznawania takie, jakie mają np. antyciała w systemie immunologicznym, oparte są o prostą zasadę zamka i klucza. Te naturalnie powstałe mechanizmy nie mają zdolności przewidywania. Złożone są z wielu elementów, z których każdy spełniać może kilka różnych funkcji. Wzorowanie rozumowań na strukturze budowy mechanizmów, które są artefaktami i gdzie każdy z elementów pełni jedną funkcję, spowodowało zdaniem Dennetta, trudności w zrozumieniu działania mózgu. Na podstawie własnych wstępnych rozważań Dennett proponuje poczynienie następujących założeń[33]:

[33] D.C. Dennett, *Consciousness Explained*, s. 176.

1. Są powody by rozpoznawać.
2. Gdzie są powody, są punkty widzenia, z których się rozpoznaje lub oszacowuje.
3. Każdy podmiot musi odróżnić „tutaj wewnątrz" od „świata zewnętrznego".
4. Wszystkie rozpoznawania muszą być całkowicie spełniane przez nieokreślenie wielka liczbę „ślepych, mechanicznych" procedur.
5. We wnętrzu chronionych granic nie zawsze potrzebny jest Nadrzędny Wykonawca lub Główny Dowodzący.
6. W naturze ładne jest tym, czym ładne jest, bez względu na jego pochodzenie.
7. W naturze elementy często spełniają wiele funkcji, w ekonomii pojedynczego organizmu.

Punkt widzenia świadomego obserwatora zdarzeń jest daleką pochodną tych założeń, które zostały przyjęte na podstawie faktów przypisanych życiu prostych samoreproduktorów. Wykształcone zostały z nich organizmy posiadające mózgi, w których zapoczątkowany został proces nazywany przez Dennetta wytwarzaniem [*produce*] przyszłości. Każdy podmiot musi nieustannie rozważać, w jaki sposób powinien postępować, do czego niezbędny jest układ nerwowy. Mózgi nazywa Dennett maszynami przewidującymi [*anticipation machines*] mającymi za zadanie wytworzenie odpowiedniej reakcji na sygnał pochodzący ze środowiska. Proste systemy nerwowe czerpią ze środowiska jedynie niewielką ilość informacji i mogą tylko reagować bezpośrednio na aktualne zdarzenia. Systemy bardziej złożone odnajdują prawa lub przybliżone prawa rządzące światem i wykorzystując je wytwarzają przyszłość służącą wzmocnieniu swoich granic.

Umysły ludzkie wyposażone są w wiele własności, odziedziczonych na drodze ewolucji po prostych systemach nerwowych. Dennett wykorzystuje to dla pokazania, jak ukształtowało się postrzeganie świata przez człowieka i uświadamianie co jest postrzegane. Pisze on, że w przeciwieństwie do prostych organizmów reagujących na bezpośrednie bodźce, wiele ssaków zaczęło gromadzić informacje dla nich samych. Informacje te nie miały bezpośredniego wpływu na organizm w momencie ich gromadzenia, lecz były rozpatrywane pod względem ich użyteczności w przyszłości dla zmiany najbliższego środowiska organizmu na bardziej mu przyjazne. W tym celu, głównie naczelne, wykształciły bardzo ruchliwe oczy pozwalające właściwie bez przerw obserwować otoczenie. Te zmiany ewolucyjne związane były, zdaniem Dennetta, z pojawieniem się ciekawości lub głodu epistemicznego. Dla jego zrealizowania organizmy nie wykształciły zupełnie nowych systemów, lecz wykorzystały istniejące już systemy przechowywania informacji. U ssaków nastąpiło wyspecjalizowanie się danych obszarów mózgu półkul: prawej i lewej. Dennett zauważa, że szczególne znaczenie miało powstanie indywidualnych fenotypów, niezdeterminowanych, co do swego zachowania w momencie urodzenia. Obdarzonych plastycznością systemu nerwowego umożliwiającą uczenie się w czasie życia pojedynczego organizmu. Plastyczność i głód epistemiczny przyspieszyły proces ewolucji.

Niektóre z warunków środowiskowych nie mogą spotykać się ze stereotypową reakcją organizmu, taką jak sen i czuwanie, czy przystosowanie do kolejnych pór roku. Sytuacjom nieprzewidywalnym organizm musi sprostać ucząc się nowych zachowań i tworząc nowe projekty w obrębie właściwej mu plastyczności, co Dennett nazywa pourodzeniowym ustalaniem projektów[34]. Obejmuje on tym terminem zarówno uczenie się koncentrowania wzroku na danym punkcie, jak i naukę mechaniki kwantowej. Ustalanie projektów w czasie życia organizmu odbywa się w sposób wysoce podobny do ustalania ich przed urodzeniem,

[34] Tamże, s. 183.

czyli na drodze doboru naturalnego. Projekty ustalone ewolucyjnie i gotowe już w momencie narodzin pełnią rolę mechanicznego selektora, który wybiera spośród wielu wygenerowanych możliwości. Na tej zasadzie oparty jest behawioryzm Skinnera, używający par bodziec-reakcja jako propozycji, spośród których mechanizm wybiera właściwe, co potwierdzane jest nagradzaniem. Jednak ten rodzaj warunkowania operacyjnego został powszechnie uznany za zbyt wielkie uproszczenia. Pourodzeniowe ustalanie projektów w umysłach ludzkich jest bardziej złożone.

Dennett proponuje jedną z wielu istniejących wersji przeniesienia procesu ewolucyjnego do wnętrza mózgu. Plastyczność mózgu określa jako jego zdolność do reorganizacji pod wpływem bodźców pochodzących ze środowiska. Proces ten jest procesem mechanicznym, analogicznym do zasady działania ewolucji. Przeprojektowywane w indywidualnych mózgach są różne struktury kontrolne lub mające wpływ na zachowanie organizmu. Wybór nowego projektu dokonywany jest przy użyciu pewnego mechanicznego procesu, który jest genetycznie zainstalowany w systemie nerwowym. Zdolność dokonywania takich wyborów i samoprojektowania się zapewnia, obdarzonym tą zdolnością organizmom, przewagę nad innymi i przyspiesza ich ewolucję. Fenomen ten znany jest jako efekt Baldwina[35].

Dennett rozważa teoretycznie populacje pewnego gatunku, którego przedstawiciele nie mają plastycznych mózgów. W takiej populacji wrodzone projekty właściwe są rzadkością, a ich posiadacze mają małą szansę na przetrwanie, gdyż są nieliczni. Inaczej sytuacja wygląda dla populacji, której przedstawiciele posiadają plastyczne mózgi. Uczą się oni właściwych posunięć i nie ma wśród nich wielkiej różnicy pomiędzy małą grupą posiadaczy projektów właściwych, a większością wyposażoną w projekty nieprzydatne. Wciąż jednak największą szansę na przetrwanie mają te jednostki, których wrodzone projekty są najbardziej zbliżone do projektów pożądanych. Gwarantuje to przekazywanie kolejnym pokoleniom coraz lepszego wyposażenia genetycznego i stwarzanie coraz lepszych warunków początkowych indywidualnego uczenia się. Nic, czego osobniki uczą się w czasie swego indywidualnego życia nie zostaje przekazane genetycznie. Postęp polega na tym, że w przekazywaniu wyposażenia genetycznego uczestniczą osobniki o najbardziej przydatnych projektach wrodzonych.

Mózg, czyli przewidująca maszyna, dostosowuje się do warunków otoczenia w kilka milisekund i wytwarza przyszłość zgodnie z własnym planem. Maszyna ta składa się z wielu systemów i podsystemów wykształconych na drodze ewolucji z prostszych układów nerwowych, które miały do czynienia z zalewem różnorodnych informacji. Doprowadziło to do powstania problemu przechowywania i organizowania tych informacji oraz ewentualnego udziału w tym wyższych systemów kontroli. Systemy takie, pisze Dennett, nie zawsze są konieczne. Przykładowo system immunologiczny organizmu może działać bez kontroli ze strony systemów wyższych rzędów. Mózg człowieka wyewoluował przez okres około sześciu milionów lat, z mózgu wspólnego przodka ludzi i szympansów. Jest teraz od mózgu tego przodka czterokrotnie większy. Trzy i pół miliona lat temu przodek człowieka poruszał się już na dwóch nogach, lecz miał jeszcze mózg wielkości mózgu małpy. Około dwóch i pół miliona lat temu, kiedy nastąpiła epoka lodowcowa, rozpoczęła się, tak zwana przez Dennetta, wielka encefalizacja mózgu, której najważniejsza część dobiegła końca 150,000 lat temu. W skali ewolucji rozwój ten jest bardzo szybki, przypomina eksplozję. Natomiast mózg wczesnego *homo sapiens*, który żył od około 150,000 do 10,000 lat temu jest prawie nieodróżnialny od mózgu człowieka, pod względem rozmiarów i kształtu.

[35] Tamże, s. 184.

Rozwój mózgu nastąpił więc przed pojawieniem się języka i język nie może być za ten rozwój odpowiedzialny. Język i inne zdobycze mentalne gatunku ludzkiego to, jak pisze Dennett używając metafory komputerowej, *software*, pracujący na *hardware*, którym jest przygotowany wcześniej przez ewolucję mózg.

Projekty, które rozważa Dennett, zarówno wrodzone jak i wyuczone, są w pewnym sensie zależne środowiskowo i w tym wąskim sensie można, zdaniem Dennetta, nazywać je reprezentacjami. Zastanawia się on też, jakie cechy projektów mogą upoważniać do rozważania ich jako reprezentacji. Coś w mózgu musi ulec zmianie, aby można było mówić, że zmiana w środowisku zewnętrznym pozostawiła tam jakiś ślad. Zdolność do tworzenia reprezentacji posiada wiele niższych zwierząt, a w człowieku zdolność ta rozwinęła się ze zwielokrotniona siłą. Stwierdza, że dzieje się tak prawdopodobnie dzięki procesowi generowania i testowania, a następnie wybierania określonych wzorów aktywności neuronalnych w korze mózgowej. Ważny jest tutaj rozwój kultury i zdolności wzajemnego przekazywania sobie jej wytworów. Dennett uważa to za kolejne medium ewolucji. Przedstawia on eksperyment myślowy, w którym proponuje wyjaśnienie rozwinięcia zdolności do tworzenia reprezentacji, zdolności językowych i myślenia u hominidów poprzez zadawanie przez indywiduum prostych protopytań, przy braku współplemieńca, który mógłby udzielić na nie odpowiedzi. Wtedy umysł poddaje się autostymulacji i generuje projekty próbne, spośród których wybiera właściwą odpowiedź. Podobnie jak z pytaniami sytuacja wygląda z prostymi rysunkami i figurami przestrzennymi. Pomagają one w generowaniu projektów i mogą być wytwarzane w samym tylko umyśle.

Różnego rodzaju znaki i symbole obecne są w kulturze i z niej czerpane pomagają indywiduom w generowaniu pożądanych projektów. Sposób, w jaki kultura stała się medium przekazu projektów i wzorców jest, zdaniem Dennetta, istotny dla zrozumienia źródeł świadomości. Jednym z pierwszych posunięć złożonego procesu samoprojektowania jest nabycie języka. Odbywa się to z taką łatwością, że nie ma właściwie wątpliwości, co do istnienia w genotypie człowieka specjalnego wzorca nabywania języka. Zdaniem Dennetta w języku pojawiają się reproduktory działające na takiej samej zasadzie, jak reproduktory cząsteczek DNA. Reproduktorami tymi są idee, lecz nie idee proste w sensie, w jakim pisali o nich empiryści brytyjscy, lecz idee złożone, takie jak np. idea koła, alfabetu, rachunku różniczkowego, czy ewolucji na drodze doboru naturalnego. Takie istniejące w kulturze cząstki pojęciowe nazywa Dennett memami [*memes*]. Nazwa memy pochodzi od greckiego *mimesis* oznaczającego imitację, a wprowadził ją zoolog R. Dawkins[36]. Memy są najmniejszymi cząstkami, które mogą się reprodukować zachowując swoją własną specyfikę. Są one jednostkami przekazu kulturowego lub imitowania. W *The Selfish Gene* Dawkins pisał o ewolucji memów kultury[37]. Ewolucję tę rozumiał w sensie dosłownym, czyli jako proces doboru naturalnego. Memy kultury i geny są po prostu różnymi rodzajami replikatorów. Reprodukują się one w różnych mediach przekazu i z różną prędkością, lecz podlegają temu samemu prawu doboru naturalnego.

Jest to, zdaniem Dennetta, właściwy sposób myślenia o ideach, choć wywołuje on problemy świadomości i jaźni. Mają one być rozumiane jako wytwory, zarówno procesu doboru naturalnego, jak i ewolucji kulturowej. Zasadą działania memów, tak jak i genów, jest reprodukowanie mające zapewnić przetrwanie. Pewne memy, jak np. współpraca, edukacja, redukcja zbrojeń są pożądane z ludzkiego punktu widzenia, lecz funkcjonują społecznie również memy kontrowersyjne oraz szkodliwe.

[36] Tamże, s. 201.
[37] Tamże, s. 202.

Zarówno geny jak i memy są niewidoczne. Geny przenoszone są w łańcuchach DNA, a memy w nośnikach kultury, którymi są książki, obrazy, narzędzia, przysłowia itp. Zniszczenie wszystkich fizycznych nośników memów jest równoznaczne z ich eksterminacją. Dennett zastanawia się nad kryterium wyboru memów, które zostaną zachowane i będą reprodukowane. Pisze on, że choć zwykle uważa się za takie memy prawdy, dobra i piękna, to przekazywane są też kulturowo memy o wartościach przeciwnych. Świadczy to, jego zdaniem, o tym, że kryterium wyboru danego memu jest po prostu jego bycie dobrym reproduktorem. Memy, jak i geny są potencjalnie nieśmiertelne, lecz zależne od istnienia nośników fizykalnych, które są zniszczalne. Przetrwanie memów nie zależy od trwałości ich nośników, lecz od ich zdolności do reprodukcji. Umysły mają ograniczoną zdolność przyjmowania i kumulowania memów. Jest to, jak zauważa Dennett, czynnikiem warunkującym wybór określonych memów ze sfery ich występowania. Własnością memów jest kształtowanie takiego fenotypu, który minimalizuje i usuwa warunki środowiska niesprzyjające reprodukcji danego memu. Umysł ludzki jest artefaktem ukształtowanym przez memy kultury strukturyzujące ludzki mózg. Mózgi różnią się mikrostrukturami wytworzonymi przez te, w nich odwzorowane, najmniejsze cząstki kulturowej reprodukcji idei. Problem kształtowania się świadomości i jaźni zależy więc od tego jakie memy kultury zostają przyjęte do umysłu i w jaki sposób zostają one zorganizowane. Ewolucja mimetyczna jest, wraz z ewolucja genetyczną i plastycznością fenotypową, czynnikiem rozwoju świadomości i jaźni człowieka. Dennett podaje hipotezę, której broni:

> Świadomość ludzka jest sama przez się wielkim kompleksem memów (lub dokładniej, efektem [działania] memów w mózgu), które mogą być rozumiane jako operacje wirtualnej maszyny „von Neumannowskiej” zastosowanej w równoległej architekturze mózgu, która nie została zaprojektowana dla takiej aktywności. Siła oddziaływania tej maszyny wirtualnej znacznie zwiększa siłę znajdujacego się u jej podstaw organicznego *hardware*, na którym ona pracuje, ale równocześnie wiele z jej ciekawych cech, a szczególnie jej ograniczeń, może zostać wyjasnionych jako produkty uboczne tego systemu połączeń, który umożliwia zastanawiające, lecz skuteczne powtórne użycie istniejącego organu dla nowych celów.
>
> [D.C. Dennett, *Consciousness Explained*, s. 210]

Hipotezę swą przedstawia Dennett używając żargonu badaczy sztucznej inteligencji, gdyż jego zdaniem najlepiej nadaje się on do tego celu. Brytyjski matematyk Alan Turing rozwinął teoretycznie koncepcję maszyny nazywanej maszyną Turinga, a węgiersko-amerykański matematyk i fizyk John von Neumann przekształcił ten eksperyment myślowy w projekt elektronicznego komputera. Komputery, zupełnie jak mózgi, nie posiadają ustalonych do końca projektów w momencie ich zmontowania. Poza tym, że mają *hardware*, czyli odpowiednik organicznej struktury mózgu, przechowują w pamięci dane i programy, czyli *software*. Podlegaja one wpływom „środowiska” poprzez klawiaturę i inne urządzenia wejściowe oraz reagują na nie sygnałem na wyjściu. *Software* komputera składa się z reguł, a nie z ich fizykalnego nośnika i nazywany jest przez przedstawicieli nauk komputerowych maszyną wirtualną. Pewne szczegóły działania takiej maszyny wirtualnej są istotne dla koncepcji Dennetta.

Turing zadawał sobie pytanie, co robi jego umysł, kiedy rozwiązuje problem lub dokonuje komputacji. Następnie rozłożył procedurę działania umysłu na najmniejsze możliwe kroki i podał ich formalizacje. Jego pomysł składał się z następujących elementów[38]:

[38] Tamże, s. 212.

1. pewien powtarzalny proces (zdarzenia następujące pojedynczo)
2. w pewnej podzielonej na części przestrzeni roboczej, do której
3. zarówno dane jak i instrukcje są wprowadzone
4. z pewnej wewnętrznej i nadzwyczaj godnej zaufania pamięci
5. gdzie poddane zostają działaniu pewnego skończonego zbioru operacji podstawowych.

W pierwotnej wersji maszyny Turinga przestrzenią roboczą był skaner nakierowany na wycinek taśmy papierowej, na której znajdował się ciąg złożony z cyfr: 0 i 1. Operacjami podstawowymi były przesunięcia taśmy w lewo lub w prawo, usunięcie 0 lub 1 i wpisanie innej cyfry lub pozostawienie cyfry bez zmiany. John von Neumann zmodyfikował pomysł Turinga projektując pierwszy, znajdujący praktyczne zastosowanie, komputer cyfrowy. Dane oraz instrukcje kodowane i przechowywane są w nim w systemie binarnym. Jeden bit informacji będący jednym ciągiem zero-jedynkowym oznacza jedną operację komputera. Skończona ilość operacji przeprowadzonych przez komputer prowadzi do jednego sekwentu zero-jedynkowego, co nazywane jest efektem szyjki butelki von Neumanna. W szybkich komputerach miliony operacji tego typu przeprowadzane są w ciągu jednej sekundy. Dennett pisze, że można takie komputery uznać za elektroniczne imitacje umysłów. Nie mogą one być imitacjami mózgów, w których pojedyncze operacje są dublowane i przeprowadzane równolegle w różnych łańcuchach połączeń neuronalnych. Dennett zauważa też, że daleko stąd jeszcze do podstawowych operacji przeprowadzanych w strumieniu świadomości. Próbę rozłożenia ich na operacje atomowe, takie, jakie zachodzą w komputerze cyfrowym proponuje Dennett jako punkt wyjściowy swojej metody badawczej.

Uproszczeniem byłoby jednak uznanie, że skoro przodek człowieka musiał zacząć myśleć, to w korze mózgowej jego lewej półkuli mózgu wykształcona została von Neumannowska maszyna wirtualna. Introspekcyjne badanie własnego umysłu było wzorem projektu takiej maszyny, lecz budowa współczesnych komputerów znacznie różni się, od jej względnie prostego projektu. Poza tym równoległa architektura mózgu, w przeciwieństwie do szeregowej architektury komputera, podawana jest często jako argument przeciwko teorii sztucznej inteligencji. Dennett stwierdza, że z pewnego punktu widzenia, różnica w architekturze nie jest dla maszyny Turinga istotna. Maszyna ta może imitować każdą maszynę komputacyjną, również mającą stany równoległe, jeżeli tylko podany zostanie wystarczający opis tej maszyny. Wtedy maszyna Turinga staje się wirtualnie tą drugą maszyną. Może ona więc być również wirtualnym mózgiem. Maszyna działająca na zasadzie – jedna rzecz w jednaj chwili, zastępuje wtedy maszynę działającą na zasadzie – wiele rzeczy w jednej chwili. W przypadku zastępowania mózgu działającego na zasadzie – milion rzeczy w jednej chwili, czas przeprowadzania operacji zwiększony zostanie milion razy. Z tego powodu badacze sztucznej inteligencji skłaniają się ku stosowaniu maszyn działających równolegle.

Dennett stwierdza, że świadomy umysł ludzki jest w zasadzie maszyną wirtualną, w rodzaju maszyny von Neumannowskiej, działającą na dostarczonym przez ewolucję *hardware*. Tak jak możliwe jest, przy podaniu wystarczającego opisu, naśladowanie przez tę maszynę każdej innej maszyny, możliwe jest również naśladowanie maszyny von Neumannowskiej przez organiczny mózg. Programy zainstalowane zostają wtedy dzięki wytworzeniu miliardów połączeń międzyneuronalnych. Świadomość ludzka, zdaniem Dennetta, jest czymś analogicznym do komputerowego *software* działającego na *hardware* organicznego mózgu. Ponieważ mózgi nie używają jednego języka i każdy z nich ma swój odrębny, wewnętrzny język naturalnej maszyny, przekazywanie oprogramowania musi odbywać się społecznie, poprzez memy kultury i dzięki plastyczności mózgów, czyli ich wysokiej tolerancji formatu.

Maszyna von Neumannowska zrealizowana w organicznym mózgu, nie jest jeszcze świadoma. Korzystnym dla przetrwania było jednak, aby uczyniła się ona przedmiotem swego systemu percepcyjnego. Taką maszynę nazywa Dennett maszyną Joyce'owską, a jej wyewoluowanie uważa za kolejny krok ku pełnemu ukształtowaniu się świadomości ludzkiej. Oczywiście nie podaje Dennett kompletnego przepisu na zbudowanie maszyny Joyce'owskiej, ani pełnego wyjaśnienia fenomenu ludzkiego umysłu. Nie należy jednak tego od niego oczekiwać, gdyż nauki poznawcze i badania nad sztuczną inteligencją wciąż są w swym stadium początkowym. Celem ich jest zaprojektowanie i skonstruowanie sztucznego umysłu, którego funkcje wzorowane są na funkcjach umysłu żywego człowieka. Dennett proponuje wiele ciekawych rozwiązań i kierunków poszukiwań, w których odwołuje się zarówno do biologii jak i technologii. Jednak dopiero przyszłość, która przyniesie pogłębione badania teoretyczne i nowe technologie, mogła będzie pokazać na ile jego pomysły były słuszne.

3. Świadomość

3.1. Krytyka dualizmu w teorii świadomości

Dualizm Kartezjański traktujący umysł jako różny od mózgu, gdyż nie złożony ze zwykłej materii, lecz ze specjalnego rodzaju tworzywa nie jest obecnie, poza pewnymi wyjątkami, popularny. Stało się tak, od kiedy G. Ryle ogłosił tę koncepcję „dogmatyzmem ducha w maszynie"[39]. Dennett pisze, że obecnie przeważnie przyjmowaną koncepcją jest materializm w różnych jego odmianach. Głosi on, że istnieje tylko jeden rodzaj tworzywa, bez względu na to, czy chodzi o przedmioty fizyki, chemii lub psychologii. Umysł jest więc, w pewnym sensie, tylko mózgiem. Dennett wyjaśnia fenomen świadomości w ujęciu materialistycznym i uzasadnia swoje podejście wykazując nieskuteczność dualizmu.

Podstawowym zarzutem, który znany był już Kartezjuszowi i nigdy nie został przezwyciężony jest zarzut niemożliwości wzajemnych oddziaływań umysłu i mózgu, jeżeli składają się one z innych rodzajów tworzywa. Kartezjusz dopatrywał się miejsca, w którym takie oddziaływanie miałoby zachodzić w szyszynce. Jego pogląd nazywany jest interakcjonizmem kartezjańskim. Dennett stwierdza, że pogląd taki narusza zasadę zachowania energii. Nie zakładamy w nim znajomości tworzywa umysłu, lecz to, że nie jest on zwykłą materią. Analizując oddziaływanie umysłu na mózg stwierdzamy następnie w tym rozumieniu, że zmiany w materialnym tworzywie połączeń neuronalnych mózgu zachodzą bez udziału tradycyjnie pojętej energii, co jest sprzeczne z zasadą jej zachowania. Dla podania fizykalnego wyjaśnienia fenomenu świadomości potrzebna będzie jednak, pisze Dennett, rewolucja w fizyce polegająca na odkryciu nowych cząstek lub oddziaływań i zwiększeniu zakresu ontologii nauk fizykalnych. Dualizm wprowadza antynaukowe nastawienie w badaniach dotyczących zależności mózgu i umysłu. Zdaniem Dennetta umiejscawia to ten problem poza zasięgiem ludzkich możliwości poznawczych. Dlatego pragnie on dualizmu uniknąć[40].

---

[39] G. Ryle, *Czym jest umysł?*, s. 48.

[40] Podobnie uważa J.J. Smart. Argumentuje on za nieistnieniem podstaw filozoficznych dla przyjęcia dualizmu i krytykuje zakładanie praw psychofizycznych. Procesy mózgowe wyróżniają się stopniem komplikacji i ujemnym sprzężeniem zwrotnym, lecz są częścią fizykalnego wszechświata. Nie będą stosowały się do nich prawa dotyczące mikrocząstek, ponieważ praw tych nie stosuje się do opisu procesów zachodzących w szerszej

W badaniach mózgu, obok potraktowania go jako jeden z wielu organów ciała, pojawia się podejście wprowadzające nowy przedmiot badania. Jest nim umysł/mózg [*mind/brain*][41]. Nauki poznawcze bez względu na to, czy bliższe są psychologii, czy sztucznej inteligencji, czy badaniom neuronalnym próbują unikać problemu interakcji: mózg-umysł. Wprowadzają one zwykle rodzaj centrum, w którym następuje połączenie działania systemów peryferyjnych. W centrum tym, nazywanym przez Dennetta teatrem kartezjańskim [*Cartesian Theatre*][42], w tajemniczy sposób pojawia się świadoma myśl. Chcąc tego uniknąć Dennett proponuje następujące założenia swojej teorii świadomości[43]:

1. Żadnej cudownej tkanki. Będę próbował wyjaśnić każdą zagadkową cechę świadomości ludzkiej w ramach współczesnych nauk fizykalnych; w żadnym miejscu nie odwołam się do niewyjaśnialnych lub nieznanych sił, substancji, czy sił organicznych. Innymi słowy zamierzam sprawdzić, co można zrobić w konserwatywnie ograniczonej standardowej nauce, pozostawiając wywołanie rewolucji w materialiźmie jako ostateczność.
2. Żadnego udawanego znieczulenia. Powiedziano o behawiorystach, że udawali znieczulenie – udawali oni, że nie mają doświadczeń, o których doskonale wiemy, że podzielają je z nami. Jeżeli chcę zaprzeczyć istnieniu pewnej kontrowersyjnej cechy świadomości, to na mnie spada ciężar pokazania, że jest ona w pewnym sensie iluzoryczna.
3. Żadnego zagłębiania się w empiryczne szczegóły. Będę próbował przedstawić wszystkie fakty naukowe właściwie tak dalece, jak są one dzisiaj poznane, lecz jest bardzo kontrowersyjne, które z intrygujących odkryć przetrzymają próbę czasu. Jeżeli miałbym ograniczyć się do faktów podręcznikowych, mógłbym być niezdolny do pokazania znaczenia pewnych ostatnio dokonanych, najbardziej otwierających oczy odkryć (jeżeli rzeczywiście one nimi są) (...).

Gdzie jest świadomy umysł, tam jest też punkt widzenia. Dennett rozważa umysł, który przyjmuje ograniczony podzbiór informacji z całości informacji zawartej w środowisku. Można przyjąć, że zachodzi to w ciągłym następstwie kolejnych współrzędnych czasoprzestrzennych. Docierająca do umysłu informacja ulega zniekształceniom, które Dennett porównuje do zakrzywienia czasoprzestrzeni w pobliżu wielkich mas lub do efektu Dopplera. Po przekroczeniu granic organizmu nie można rozważać informacji w ten sam sposób, w jaki rozważa się ją, kiedy zawarta jest w środowisku zewnętrznym. W mózgu nie ma jednego punktu, do którego ona dociera. Dennett zaznacza, że w mikroskali mózgu dla zrozumienia zjawiska istotne są przyjmowane wielkości. Czas trwania następujących tam zdarzeń mierzony jest w milisekundach. Przykładowo wypowiedzenie czterech lub pięciu sylab zajmuje 1000 ms, sygnał bólu z palca do mózgu jednym z włókien nerwowych pokrytych mieliną dociera w 20ms, a cykl neuronu wynosi 10ms.

Wiadomo, pisze Dennett, że mózg przetwarza bodźce wizualne wolniej niż bodźce słuchowe. Doświadczenia z szybkością reakcji na bodźce wzrokowe i słuchowe pokazały, że mózg pobudzany bodźcem pochodzącym z organu słuchu wysyła w odpowiedzi sygnał prowadzący do reakcji na ten bodziec o 30~40 milisekund szybciej niż przy pobudzeniu bodźcem pochodzącym z organu wzroku. Doświadczanie dźwięku lub obrazu musi następować w

---

skali. Smart jest zdania, że procesy mózgowe należy opisywać stosownymi do ich skali prawami fizykalnymi, które będą dotyczyły również doznań i doświadczeń podmiotu. Wyznaje on teorię identyczności.

[41] Por. D.C. Dennett, *Consciousness Explained*, s. 38.

[42] Tamże, s. 39.

[43] Tamże, s. 40-41.

czasie pomiędzy momentem dotarcia do narządu zmysłu a wystąpieniem reakcji. W koncepcji nazywanej przez Dennetta materializmem kartezjańskim[44], następowało ono w szyszynce, gdzie sygnały fizykalne w niewyjaśniony sposób przekształcane były w jakości mentalne. Oprócz szyszynki proponowane były również inne miejsca w mózgu, w których miało następować takie przekształcenie. Każde z takich miejsc Dennett określa mianem teatru kartezjańskiego. Metafora ta ma obrazować umiejscowienie świadomego doświadczenia w mózgu. Jest to obraz błędny, gdyż w mikroskali mózgu nie możemy brać pod uwagę równoczesności lub kolejności sygnałów docierających do widza obserwującego scenę teatru. Nie ma w mózgu jednego punktu, punktu widzenia tego właśnie widza, do którego zmierzałyby wszystkie impulsy. Ich kolejność jest względna i zależna od miejsca w mózgu, które będziemy rozważać. Teatr kartezjański należy więc zastąpić innym modelem. Dennett wprowadza model szkiców złożonych [*Multiple Drafts*][45].

W modelu szkiców złożonych wszystkie rodzaje percepcji, myślenia, jakiejkolwiek aktywności mentalnej są spełniane w mózgu przez wiele równoległych procesów[46] obróbki i interpretacji sygnałów pochodzących z wejść zmysłowych. Informacja przyjmowana do systemu nerwowego podlega ciągłym „poprawkom redakcyjnym". Dennett podaje przykład ruchu gałek ocznych, który odbywa się na zasadzie około pięciu szybkich fiksacji na sekundę. Widzący odbiera jednak zmianę punktu widzenia jako płynną. Dzieje się tak, gdyż sygnał z gałki ocznej zanim dotrze do świadomości poddawany jest pewnej obróbce. O jej zachodzeniu przekonać się można również na podstawie złudzeń optycznych z widzeniem trójwymiarowym. Procesy mózgowe stwarzają tu wrażenie postrzegania trzech wymiarów przy informacji pochodzącej z dwóch dwuwymiarowych rysunków. Nasze doświadczenia są wynikiem wielu procesów interpretacyjnych, a nie tym, co aktualnie dzieje się na siatkówce oka, na skórze czy w uchu. Dennett pisze, że początkowo mamy stosunkowo surowe jednostronne reprezentacje. Następnie są one zbierane razem i poddawane interpretacji, co jest realizowane dzięki aktywności różnych części mózgu. Tak twierdzą wszystkie teorie percepcji, a nowość wprowadzona przez model szkiców złożonych polega na tym, że jakiekolwiek rozpoznanie lub rozróżnienie cech dokonywane jest tylko jeden raz. Oznacza to istnienie określonych części mózgu odpowiedzialnych za rozpoznawanie lub rozróżnianie danych cech, ale pochodząca z nich informacja nie jest przesyłana dalej do jakiejś nadrzędnej części mózgu, gdzie zachodzi ponowna interpretacja. Takiego miejsca, nazywanego przez Dennetta teatrem kartezjańskim, nie ma w modelu szkiców złożonych.

Zmiany w określonych częściach mózgu zaangażowanych w rozpoznawanie lub porównywanie cech zachodzą w określonym czasie i przestrzeni. Ich pojawienie się, pisze Dennett, nie oznacza jeszcze pojawienia się świadomości ich treści. Jest wiele nieporozumień i problemów dotyczących tego zagadnienia. Wraz z upływem czasu następują po sobie kolejne stany neuronalne będące częściami procesów w mózgu. Dennett nazywa je czymś w

[44] Tamże, s. 107.
[45] Tamże, s. 111.
[46] Krytykuje Dennetta N. Block. Jego zdaniem Dennett zastępuje tylko materializm kartezjański, modularyzmem kartezjańskim [*Cartesian Modularism*]. Różnica polega tylko na rozmieszczeniu tego wszystkiego, co miało dokonywać się w szyszynce, w kilku różnych miejscach w mózgu. Poza tym Block uważa, że model szkiców złożonych dotyczy tylko świadomości związanej z językiem i myśleniem [*access-consciousness*] i pomija całkowicie świadomość fenomenalną [*phenomenal consciousness*], która rozumiana jest jako doświadczenie i bezpośrednie doznania. Ten rodzaj świadomości nie może być, jak jego zdaniem chce Dennett, uznany za kompleks memów kultury. Można jednak bronić modelu Dennetta, zwracając uwagę na istniejący w nim podział sygnałów na te, które przekroczyły granicę centrum mowy i te, które jej nie przekroczyły. Słuszna natomiast wydaje się uwaga Blocka na temat materializmu kartezjańskiego i kartezjańskiego modularyzmu.

rodzaju strumienia narracyjnego [*narrative stream*][47]. Narracyjność tę rozumie jako zwielokrotnienie informacji pochodzącej z określonego stanu neuronalnego w innych częściach mózgu w pewnych stanach przekształconych, które nieprzerwanie następują po sobie. Nazywa to również strumieniem treściowym, w którym każda treść jest w pewien sposób odniesiona do innych zawartych w strumieniu treści. W zależności od tego, na której części strumienia zostaje skupiona uwaga, podmiot uzyskuje dostęp do innego fragmentu narracji. Najważniejsze w modelu szkiców złożonych jest to, że nie istnieje szkic ostateczny, który obejmuje sobą wszystkie strumienie cząstkowe poszczególnych narracji. Żaden strumień świadomości nie obejmuje sobą całości treści, do których możliwy jest dostęp. Narracja ma swe następstwa, wątki równoległe i przecinające się. Mogą one być ze sobą zestawiane, lecz zestawienie zupełne nie jest możliwe.

Zmiana schematu myślenia o świadomości jako teatrze kartezjańskim na schemat stosujący model szkiców złożonych jest, pisze Dennett, najtrudniejszą i jednocześnie najważniejszą częścią jego koncepcji. Dla zobrazowania tego problemu podaje przykład wrażenia ruchu powstającego przy pokazywaniu szybko po sobie następujących obrazów. Zjawisko to zostało nazwane fenomenem *phi*. Nazwę taka nadał mu Max Wertheimer[48], który zajmował się nim jako pierwszy. Dennett przytacza pytanie filozofa Nelsona Goodmana, który spytał zajmujących się fenomenem *phi* psychologów: Paula Kolersa i Michaela Grünaua, co stanie się, jeżeli dwa następujące po sobie obrazy będą różnych kolorów. W przeprowadzonym przez psychologów eksperymencie obrazy były eksponowane przez 150 milisekund każdy, z przerwą trwającą 50 milisekund. Okazało się, że widoczna na obrazach figura zmieniała kolor w połowie trwania swego płynnego ruchu. Ruch ten był tylko wynikiem obróbki sygnałów wejściowych w systemie percepcyjnym, a Goodman zastanawiał się jak jest możliwe, że w polu percepcji wizualnej wypełniony zostaje czas przerwy pomiędzy ekspozycjami. Dennett wyjaśnia ten fenomen przypisując mózgowi posiadanie informacji o kolorze figury zanim informacja ta zostanie uświadomiona.

Informacja zawarta we wzorze odbitym na siatkówce oka jest następnie rozprowadzana po różnych częściach mózgu. Luka w polu widzenia zostaje wypełniona, gdyż świadome doświadczenie zmieniającego kolor, ruchomego obrazu następuje później niż, jak pisze Dennett, prawdopodobnie nieświadome doświadczenia obrazów pojedynczych i nieruchomych. Pochodząca z tych obrazów informacja zostaje użyta do wypełnienia luki w polu widzenia. Gwarantuje to szybkość przebiegu impulsów neuronalnych i opóźnienie uświadomienia w stosunku do momentu odbicia wzoru na siatkówce oka. Czas ten potrzebny jest dla zajścia odpowiednich procesów neuronalnych potrzebnych dla stworzenia iluzji płynnego ruchu. Dennett nie wchodzi w szczegóły dotyczące tych procesów, ale zaznacza, że nie ma takich zamiarów. Jego celem jest tylko stworzenie ogólnego zarysu teorii, której uzupełnienie będzie mogło nastąpić wraz z rozwojem nauk poznawczych i teorii sztucznej inteligencji.

Rozważając pamięć Dennett wprowadza podoświadczeniową orwelliańską zmianę wspomnień i stalinowskie zeznania dotyczące sfałszowanych dowodów[49]. W powieści G. Orwella pod tytułem *1984* wspomnienia ludzkie mogły być zmieniane przez Ministerstwo Prawdy, a praktyką reżimu stalinowskiego było przedstawianie sfałszowanych dowodów dla ukrycia prawdy. Różnica pomiędzy tymi dwoma praktykami polega na wprowadzaniu fałszywych wspomnień na temat zdarzeń, które nigdy nie zaistniały w pierwszym przypadku i

[47] Por. D.C. Dennett, *Consciousness Explained*, s. 113..
[48] Tamże, s. 114.
[49] Tamże, s. 116-117.

zachowywaniu wspomnień prawdziwych na temat sfałszowanych zdarzeń w drugim. Rozpoznanie, która z tych metod zafałszowywania pamięci została użyta, jeżeli sfałszowane zostały wszystkie wspomnienia i artefakty, jest zdaniem Dennetta niemożliwe. Przenosi on następnie te rozważania z makroskopowej skali świata do skali pojedynczego umysłu. Jakiekolwiek wspomnienia pojawiające się po danym doświadczeniu empirycznym mogą zanieczyszczać jego przypominanie. Działają one wtedy jak zmiana orwelliańska. Nie można mieć pewności, od jakiego momentu i w jakim stopniu zanieczyszczają one przypominanie danego zdarzenia. Mogą one też wprowadzać zanieczyszczenia już na samym wstępie procesu przetwarzania informacji i mamy wtedy do czynienia za stalinowskim fałszowaniem zdarzeń. W pierwszym wypadku wspomnienia podmiotu są iluzją, w drugim cierpi on na halucynacje.

Dennett zauważa, że rozróżnienie takie traci swą siłę, gdy zwróci się uwagę na czasowe i przestrzenne rozmazanie wspomnień. Wspomnienia mogą wzajemnie się wywoływać i łączyć. W tradycyjnym materializmie kartezjańskim rozwiązaniem problemu prawdziwości wspomnień było ustanowienie pewnego stanu końcowego, do którego one docierają i ich porządku czasowego. Dennett stwierdza jednak, że takiego stanu nie ma, a porządek czasowy nie może być użyty dla porządkowania wspomnień danych doświadczeń w pamięci. Alternatywne rozwiązanie przedstawia Dennett na przykładzie fenomenu *phi*.

W mózgu może funkcjonować mechanizm stalinowski, który uzupełnia luki pomiędzy poszczególnymi doznaniami wzrokowymi tak, że uzyskują one ciągłość. Zanim zostaną uświadomione są w pewnym sensie sfałszowane, przez dodanie potrzebnych dla uzyskania ciągłości doznań nieprawdziwych. Może jednak również funkcjonować w mózgu mechanizm orwelliański, który stosowany jest już po uświadomieniu sobie doznań wzrokowych. Dennett podkreśla, że założenie funkcjonowania obu tych mechanizmów jest zgodne z tym, co podmiot myśli, mówi i pamięta. Jego zdaniem są one jednym i tym samym mechanizmem. Jedyna różnica pomiędzy nimi dotyczy miejsca i czasu, w którym następuje wypełnienie luki w doświadczeniu wizualnym. Stwierdzenie gdzie to rzeczywiście następuje jest niemożliwe i Dennett proponuje hipotezę braku określonego miejsca. Postuluje on również brak ostrej linii oddzielającej świadomość od nieświadomości.

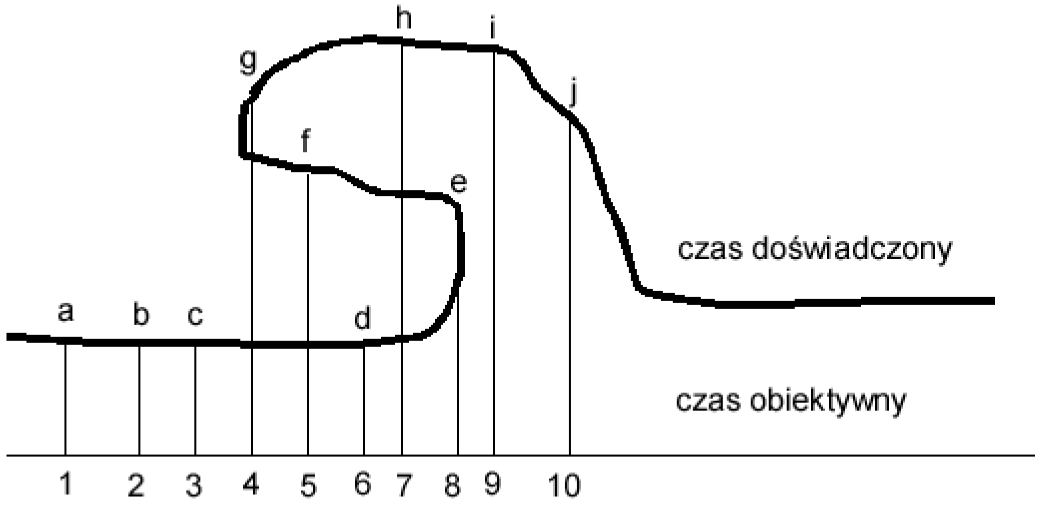
Powszechną opinią, którą przytacza Dennett jest, że fenomeny doświadczenia wizualnego są dzięki określonym funkcjom mózgu rzutowane [*projected*]. Nie są one jednak jego zdaniem rzutowane w zewnętrzną przestrzeń fizyczną, a w przestrzeń fenomenalną [*phenomenal space*]. Porównuje on ją do przestrzeni logicznej i przestrzeni świata fikcji literackiej. Jej własności ustanawiane są w perspektywie rozpatrywanego heterofenomenologicznie podmiotu. Istotne jest więc jakimi rzeczy wydają się dla podmiotu, a nie jakimi one rzeczywiście są. W modelu szkiców złożonych nie ma rzeczywistości doświadczenia niezależnego od wpływu nośników treści na pamięć i działanie. Teatr kartezjański zachowuje rozróżnienie na rzeczywistość i pozory, ale to wytwarza kategorię, jak to nazywa Dennett, rzeczy obiektywnie subiektywnych. Podmiot przypisuje im obiektywnie ważne własności, choć są one obecne tylko w jego subiektywnym przedstawieniu. Model szkiców złożonych natomiast można nazwać operacjonalizmem w pierwszej osobie [*first person operationalism*][50]. Odrzuca on możliwość uświadomienia bodźca bez wystąpienia przekonania co do jego wystąpienia. Zanegowanie teatru kartezjańskiego jest równoznaczne z zanegowaniem kartezjańskiego dualizmu.

[50] Tamże, s. 132.

Nasuwa się tutaj pytanie związane z omawianym przez Dennetta rzutowaniem fenomenów. Czy nie powraca tu teatr kartezjański? Nie można oczywiście mówić o kartezjańskim dualizmie, ale krytyka teatru kartezjańskiego mogłaby być tylko przeniesieniem go na grunt sprzężeń zwrotnych określonych modułów mózgu biorących udział w procesie percepcyjnym, z aferentnymi zakończeniami organów zmysłów[51]. Dennett omawia podobieństwo pomiędzy wyobrażaniem a widzeniem. Wydaje się, że można by spróbować pogłębić badania w tym kierunku. Widzenie lub wyobrażanie jest, mówiąc metaforycznie, sceną fizykalnego teatru, a widzem byłaby tutaj jaźń.

Dennett przedstawia działanie modelu szkiców złożonych na przykładzie doświadczenia wzrokowego. Pobudzenie wzrokowe wywołuje łańcuchy zdarzeń w korze mózgowej. Spełniają one coraz bardziej wyspecjalizowane funkcje, ale nie ma w mózgu miejsca, które łańcuchy te musiałyby przekroczyć, by doświadczenie wzrokowe zostało uświadomione. Strumień następujących po sobie i równolegle zdarzeń neuronalnych wywołuje różne skutki i reakcje organizmu. Jedną z nich jest uświadamianie. Jego efektem jest strumień świadomości, w którym uświadomione zdarzenia niekoniecznie muszą występować w porządku czasowym zgodnym z czasem obiektywnym. Dennett zamieszcza wykres, na którym subiektywna kolejność występowania doznań jest inna niż obiektywna (rys. 1).

Nie występuje tu ani czyste orweliańskie fałszowanie wspomnień, ani czyste stalinowskie fałszowanie doświadczenia. Nie ma kolejności fałszowania wspomnień, gdyż kolejność ich zostaje ustanowiona jeszcze w procesie ich formowania. Nie ma też fałszowania doświadczeń, gdyż doświadczenie podmiotu może być tylko subiektywne. Dennett odrzuca kategorię rzeczy obiektywnych, występujących w doświadczeniu pojedynczego podmiotu.



Rysunek 1.

[51] Podobnie krytykowałby Dennetta. N. Block, wprowadzając modularyzm kartezjański [*Cartesian Modularism*]. Zamiast szyszynki mielibyśmy wiele różnych miejsc w mózgu, istotnych ze względu na wytworzenie świadomości.

3.2. Struktura świadomego umysłu

Zamiast pojedynczego strumienia świadomości Dennett wprowadza wiele kanałów odpowiedzialnych za poszczególne funkcje i opisywanych za pomocą modelu szkiców złożonych. Składają się one na maszynę wirtualną wytworzoną na drodze wspólnego działania wielu modułów funkcjonalnych. Niektóre z nich człowiek dzieli z innymi gatunkami zwierząt, a część z nich została wysoce wyspecjalizowana, głównie dzięki nabyciu zdolności językowych i przekazywaniu memów kultury. Było to możliwe również dzięki bezsłownym wyobrażeniom i innym strukturom danych, które zajmują miejsce w mózgu. Dennett przytacza hipotezę Francisa Cricka i Christofa Kocha[52], którzy zasugerowali, że jedną z funkcji świadomości jest prezentowanie rezultatu komputacji, co odbywa się dzięki zsynchronizowaniu częstotliwości oscylacji zakończeń neuronów na 40Hz. Hipoteza ta nie podaje jednak, komu wynik komputacji zostaje zaprezentowany, co jest problemem teorii świadomości korzystających z osiągnięć nauk neuronalnych. Unikają tego problemu również modele umysłu proponowane przez psychologię poznawczą i teorię sztucznej inteligencji. Używają one pamięci roboczej lub operacyjnej, która zastępuje teatr kartezjański, ale na ogół nie określają gdzie mieści się ona w mózgu. Dlatego, pisze Dennett, wbrew swoim zamiarom badawczym przedstawiciele nauk neuronalnych zostają dualistami, a psychologowie poznawczy mówią o automatach lub o zombi, czyli istotach pozbawionych świadomości.

Dennett proponuje teorię świadomości, która korzysta z osiągnięć wspomnianych nauk, lecz używa ich w nowy sposób. Jego zdaniem błędem jest zakładanie przez pewne badania prowadzone w obrębie nauk neuronalnych, istnienia w mózgu określonego miejsca, gdzie rezultaty wszystkich poszczególnych procesów zostają zestawiane razem[53].

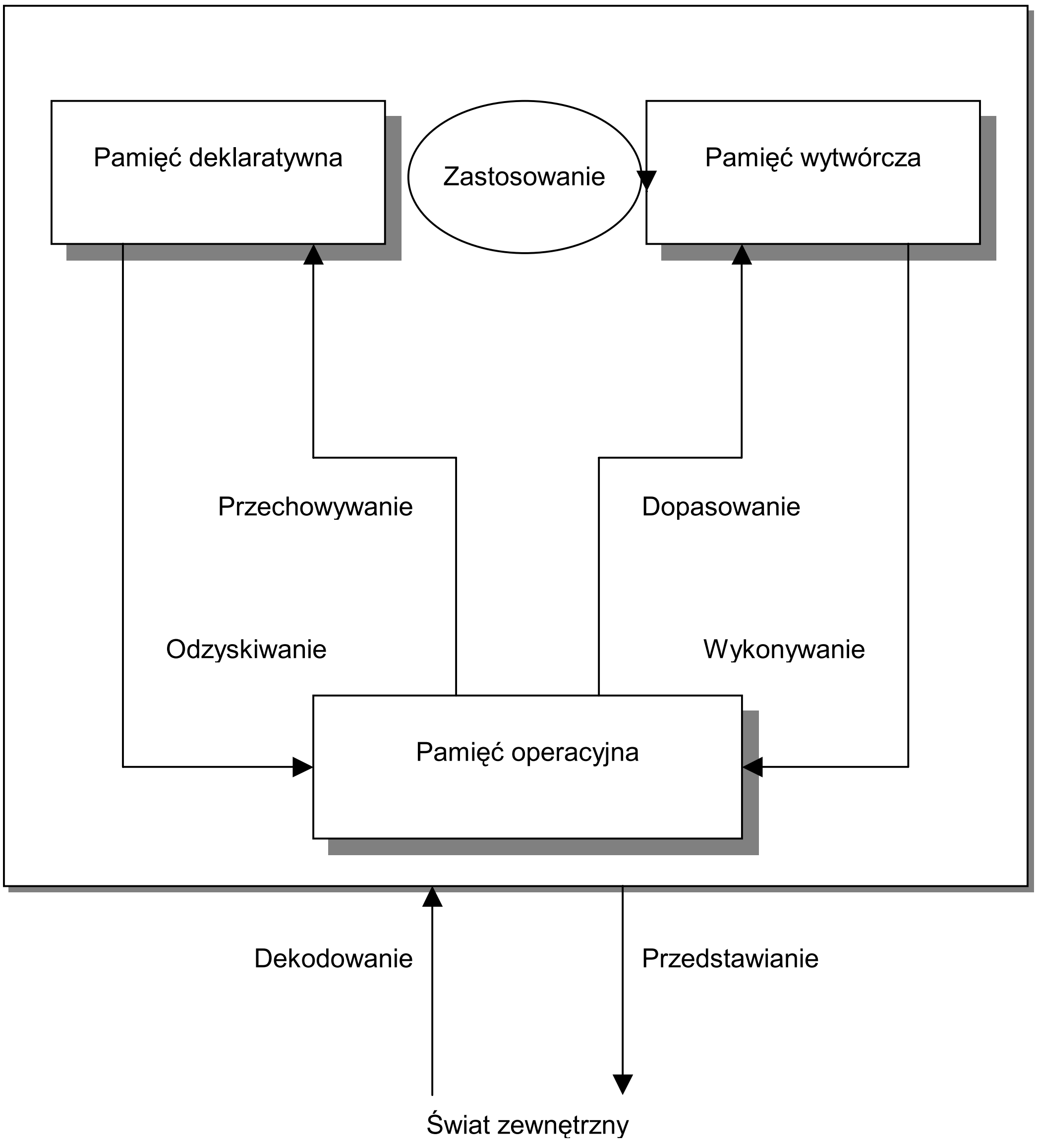
W teorii J. Fodora poszczególne moduły umysłu (bez względu na umiejscowienie ich w mózgu) kończą swe działanie w miejscu, gdzie przebiega linia oddzielająca od nich jednostkę centralną wyznaczania przekonań [*belief fixation*][54]. Zachodzi w niej globalny, nie podzielony na moduły proces. Dennett krytykuje takie podejście pisząc, że żadna z gałęzi nauk kognitywnych nie wyjaśnia, jak taka jednostka centralna mogłaby działać. Homunculusy, demony, czy podmioty [*agents*], których używa teoria sztucznej inteligencji, nauki komputerowe i psychologia poznawcza, są tylko jednostkami o określonych kompetencjach. Dennett pisze, że jednostki te stosowane są przez wszystkie odmiany funkcjonalizmu. Dla wyjaśnienia ich współpracy proponuje on model myślenia sekwencyjnego [*sequential thinking*][55].

W modelach sztucznej inteligencji na bazie funkcji podstawowych budowane są modele bardziej złożone, bliższe realistycznym modelom funkcji poznawczych. Szyjka butelki von neumannowskiej rozciągnięta zostaje do rozmiarów przestrzeni roboczej lub pamięci operacyjnej. Zaprojektowane zostają również pewne modele podstawowych funkcji psychologicznych. Mają one stosunkowo elastyczne sposoby przyjmowania instrukcji i działania, w przeciwieństwie do sztywnego zorientowania na instrukcje w architekturze von neumannowskiej. Podstawową zdolnością komputerów jest warunkowa reakcja typu: JEŻELI-TO [*IF-THEN*], od której Dennett zaczyna projektowanie architektury umysłu.

[52] Por. D.C. Dennett, *Consciousness Explained*, s. 255.
[53] Przeciwny temu jest również J. Fodor. Uważa on, że typy stanów fizycznych nie muszą odpowiadać typom stanów psychicznych. Założenie to związane jest z teorią ekwipotencjalności Lashey'a. Zgodnie z nią wszystkie funkcje psychiczne mogą być spełniane przez różne struktury mózgu.
[54] Por. D.C. Dennett, *Consciousness Explained*, s. 260.
[55] Tamże, s. 263.

Podaje on przykład systemu wytwarzania [*production system*] nazywanego ACT, [*act star*][56], (rys. 2).

W systemie takim Dennett rozważa takie sytuacje JEŻELI-TO, kiedy przedstawione może zostać więcej niż jedno rozwiązanie. Nie wszystko, co taki system wie, może dostać się do pamięci roboczej jednocześnie, więc należy określić na jakiej zasadzie zawartość tej pamięci zostaje określana. Jest to stary platoński problem gołębnika[57], z którego przywołać należy właściwego ptaka we właściwym czasie.



Rysunek 2.

[56] Tamże, s. 265.
[57] Por. Platon, *Teajtet*, s. 122-123.

Dla systemu ACT Dennett podaje pięć zasad, na podstawie których dokonany zostaje wybór właściwej zawartości pamięci roboczej, czyli przywołanie właściwego ptaka[58]:

1. Stopień dopasowania: jeżeli pewien wytwór klasy JEŻELI jest w jakiś sposób lepiej dopasowany niż inne, to ma pierwszeństwo.
2. Siła wytwórcza: wytwory, które były ostatnio skuteczne mają większą, skojarzona z nimi „siłę", która daje im pierwszeństwo w stosunku do wytworów o mniejszej „sile".
3. Strumieniowość danych: ten sam wytwór nie może być dopasowany więcej niż jeden raz (jest tak, by zapobiec nieskończonym pętlom i podobnie, choć słabiej, korzeniom).
4. Swoistość: jeżeli dwa wytwory pasują do tych samych danych, wygrywa wytwór o bardziej swoistej klasie JEŻELI.
5. Przewaga celu: pomiędzy wytworzonymi jednostkami przechowywanymi w pamięci są cele. Może być tylko jeden aktualnie czynny cel w danym czasie w pamięci roboczej ACT i jakikolwiek wytwór, którego podanie na wyjście pasuje do czynnego celu ma pierwszeństwo.

Dennett pisze, że systemy takie jak ACT są ciągle znacznym uproszczeniem, lecz właściwy jest kierunek badań prowadzący od maszyny von neumannowskiej do takich systemów. Najlepszym sposobem sprawdzenia ich trafności jest zdaniem Dennetta, uruchomienie systemów zbudowanych na podstawie takich właśnie projektów. Największą szansę mają projekty uwzględniające możliwość ich zrealizowania w organicznym mózgu. Dennett zaznacza również, że podaje tylko szkic pasujący do całej rodziny różnych teorii świadomości.

Na bardziej szczegółowym poziomie badań należy pokazać, jak homunculusy, czy demony rozpoznawania wzorów zostają zrealizowane w sieci neuronalnej. Ostatnim osiągnięciem teorii sztucznej inteligencji, które może zbliżyć modele poznawcze do modeli neuronalnych jest koneksjonizm nazywany też PDP[59], co oznacza procesy równoległe. Jego składowe przypominają sieć neuronalną w mózgu. Opisywać ten poziom oraz łączyć go z poziomami: psychologicznym i poznawczym mogą inne teorie. Dennett pisze, że nie preferuje żadnej z nich, lecz chce wydobyć ze wszystkich znanych teorii najbardziej istotne idee i na tej podstawie określić kierunek poszukiwań właściwej teorii świadomości. Z wybranej przez niego metody badawczej wynika jednak, że pomimo deklarowanej intencji pozostania neutralnym, jest on funkcjonalistą.

Zdaniem Dennetta po przekształceniu architektury von neumannowskiej w architekturę systemów wytwarzania takich jak ACT, nastąpiło przesunięcie równowagi sił od sztywnych programów w kierunku elastycznych sposobów zachowań opartych o złożone interakcje. Nie ma tu jednak różnicy jakościowej. U podstaw tych interakcji ciągle tkwią komputacyjne maszyny von Neumanna. Obrońcy teorii sztucznej inteligencji twierdzą, że sztuczne stworzenie systemów znanych obecnie jedynie jako pracujące na organicznym *hardware* mózgu, jest tylko kwestią stopnia komplikacji. Jednym z krytyków takiej hipotezy, nadmienia Dennett, jest J.R. Searle, który uważa, że żaden koneksjonistyczny komputer nie może być prawdziwie mentalny[60].

---

[58] D.C. Dennett, *Consciousness Explained*, s. 266.
[59] Z języka angielskiego: *Parallel Distributed Processing*.
[60] Podany przez Searle'a argument chińskiego pokoju pokazuje, że komputer może spełniać złożone funkcje wymagane dla zdolności językowej, bez znajomości treści i znaczenia używanych symboli. Zewnętrznie naśladuje on zachowania mentalnego umysłu, lecz wewnętrznie jest tylko pozbawioną mentalności maszyną.

Dennett stwierdza, że zwyczajem funkcjonalistów jest rysowanie diagramów, które i on sam zamieszcza. Ostrzega jednak, że stosowanie ich może spowodować niezauważenie alternatywnych możliwości rozkładu funkcji, szczególnie funkcji wielokrotnie na siebie nałożonych. Jedynym miejscem w mózgu, w którym mogłaby być umiejscowiona pamięć robocza jest cała kora mózgowa. Dennett pisze o przestrzeni roboczej, w której właściwie wszystko może wejść w interakcję ze wszystkim. W sensie anatomicznym natomiast istnieją tylko rozprowadzane po korze mózgowej impulsy neuronalne. Problemem jest więc, jak poszczególne wyspecjalizowane demony, działające w skali całego mózgu, mogą wywoływać reakcje innych demonów, w innych częściach mózgu. Powszechnie przypuszcza się, zaznacza Dennett, że funkcjonalna tożsamość demonów związana jest z miejscem zajmowanym przez nie w sieci połączeń neuronalnych. Tylko neurony połączone bezpośrednio lub pośrednio z komórkami zwojowymi tworzącymi pień nerwu wzrokowego uczestniczą w percepcji koloru. Neurony te muszą w jakiś sposób komunikować się z neuronami odpowiedzialnymi za inne funkcje.

Proponowano hipotezy, pisze Dennett, mówiące o kluczowym znaczeniu dla świadomości pewnych obszarów mózgu. Uznano, że system aktywacyjny tworu siatkowatego [*RAS - reticular activating system*], w tworze siatkowatym [*reticular formation*], w pniu mózgu i wzgórze [*thalamus*] biorą udział w budzeniu mózgu, w szybkich reakcjach na zagrożenie i nowe sytuacje oraz w skupianiu uwagi. Natomiast płaty czołowe uczestniczą w procesach myślenia, emocjonalnych i mowy, a więc jak to nazywa Dennett, w długoterminowej kontroli. Nie wyjaśnia to jednak, jego zdaniem, mechanizmu działania świadomości. Problem polega na braku komputacyjnego modelu tego mechanizmu.

Wiele zdarzeń neuronalnych współzawodniczy z sobą w mózgu w procesie wybierania spośród nich pewnego podzbioru, który zostanie wykorzystany w cyklu autostymulacji [*self-stimulation*][61]. Zdaniem Dennetta nie uczestniczy w tym świadomość. Nie ma też ostrej linii pomiędzy tym, co świadome i tym co nieświadome. Dla wyjaśnienia fenomenu świadomości Dennett wprowadza koncepcje maszyny joyce'owskiej [*Joycean machine*]. Zaznacza też, że do jego metody poszukiwań należy określenie celu, w jakim wykorzystywana jest świadomość i następujące po nim testowanie mechanizmów, które spełniają zapewniające osiągnięcie tego celu funkcje. Jako przykład takiego mechanizmu podaje trójpoziomowy projekt badacza sztucznej inteligencji Davida Marra. Poziom najwyższy to komputacyjna analiza problemu przetwarzania informacji. Poziom środkowy jest algorytmiczną analizą procesów, dzięki którym następuje przetwarzanie informacji. Poziom najniższy stanowi sieć neuronalna zapewniająca organiczne podłoże procesu. Zaznacza jednak, że podanie przetwarzania informacji jako celu działania świadomości jest uproszczeniem. Świadomość posiada wiele cech, a niektóre z nich mogą być tylko obecnie nieprzydatnymi pozostałościami po wcześniejszych etapach rozwoju gatunkowego.

Maszyna joyce'owska ma za zadanie ustalenie komunikacji pomiędzy demonami czy homunculusami. Większość prostych powiązań może być dokonywana bez udziału świadomości, lecz znajdowanie pewnych bardziej skomplikowanych połączeń wymaga koncentrowania się na nich i stosowania mnemotechniki. Są to strategie samokontroli, które pozwalają wykorzystywać materiał percepcyjny w nowy sposób. Wytworzone zostaje dzięki nim myślenie hipotetyczne i rozszerzenie pamięci długotrwałej o wspomnienia operacji przeprowadzanych w umyśle. Przypominana treść pociąga za sobą kontekst, bez którego nie ma ona dla podmiotu sensu. Treść tego kontekstu nie musi być zawsze świadoma. W razie

[61] Por. D.C. Dennett, *Consciousness Explained*, s. 275.

potrzeby podmiot nadaje im świadomie rozumiane znaczenie. Możliwa jest świadoma lub nieświadoma aktywacja różnych treści, a więc i różnych połączeń neuronalnych. Teoria Dennetta jest holistyczna, z czym wiąże się izotropiczność kory mózgowej.

3.3. Kognitywistyczna teoria świadomości

Psychologia poznawcza unika problemu świadomości, ponieważ jak zauważa Dennett, problem świadomości wciąż nie poddaje się metodom stosowanym przez psychologów poznawczych. Proponuje on następnie szkic teorii świadomości, który ma połączyć kognitywistyczne teorie percepcji, rozwiązywania problemów i użycia języka.

Zdarzenia zewnętrzne i wewnętrzne w stosunku do podmiotu Dennett dzieli na świadome i nieświadome. Świadomość nazywa również dostępem świadomości osobowej [*access of personal consciousness*][62], przez co chce zwrócić uwagę na to, że podmiotem świadomym jest cała osoba, a nie tylko jej część.

Od dostępu świadomości osobowej Dennett odróżnia dostęp komputacyjny, polegający na istnieniu łącza przepływu informacji i dostęp publiczny, którym jest przedstawianie stopnia przebiegu przeprowadzanych operacji dla wglądu z zewnątrz. Dostęp komputacyjny nie jest bezpośrednio związany z dostępem świadomości osobowej, gdyż w mózgu zachodzi wiele procesów informacyjnych, których nie jesteśmy świadomi. Bliższy temu ostatniemu sposobowi dostępu jest dostęp publiczny, gdyż w obu tych przypadkach korzystamy z podobnych sposobów przekazu informacji, z których głównym jest użycie języka naturalnego.

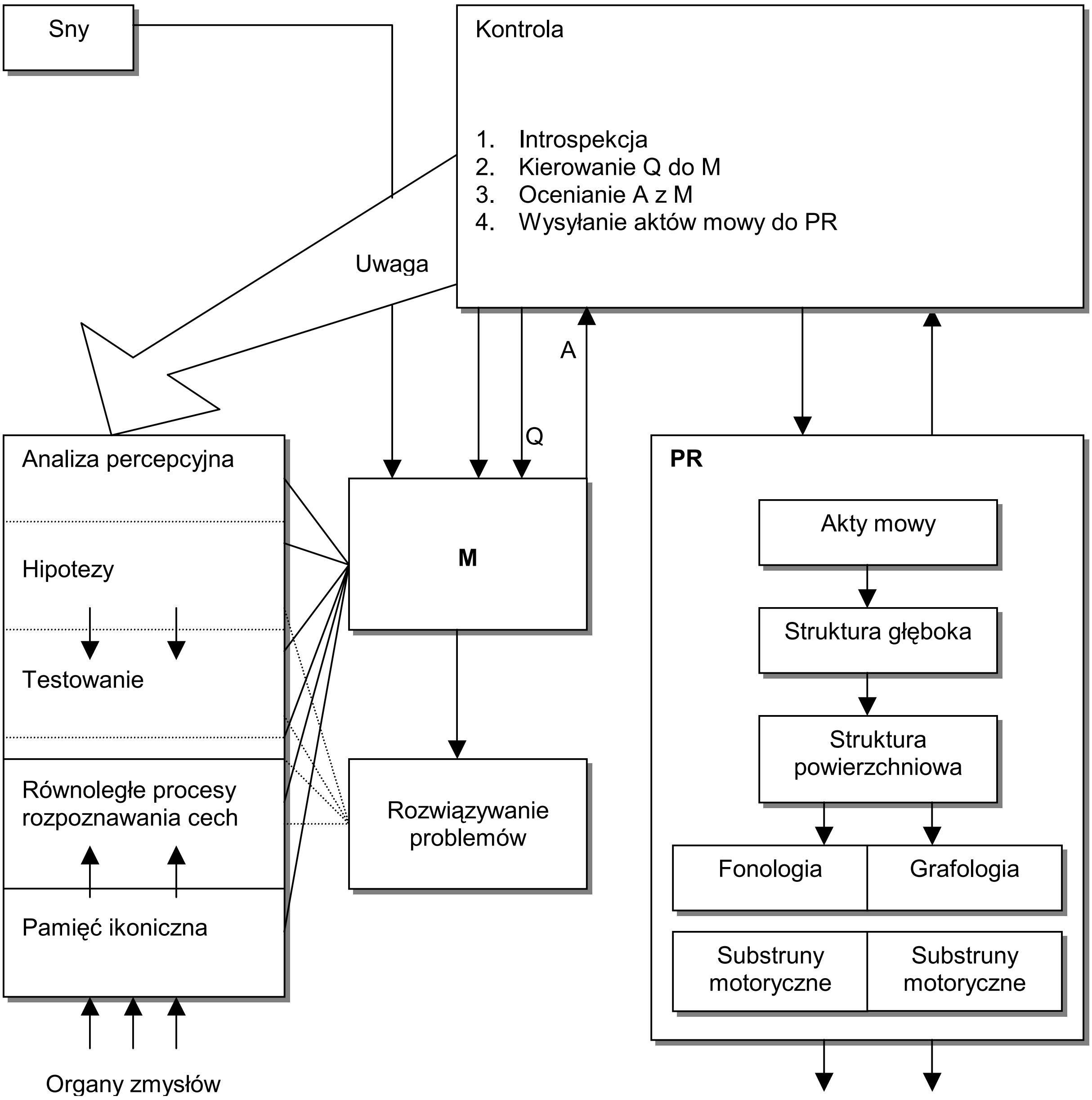
Konsekwencją uprzywilejowanej przez Dennetta funkcjonalistycznej teorii umysłu, jest przypisanie posiadania umysłów w najsłabszym, metaforycznym sensie również robotom i mechanizmom. Teorie funkcjonalistyczne nazywa Dennett teoriami poziomu subosobowego, gdyż traktują one osobę jako organizację podsystemów i próbują wyjaśnić zachowanie jako wypadkową ich współdziałania. Zdaniem Dennetta tego poziomu wyjaśniania nie można uniknąć również w teoriach psychologicznych, które korzystają z możliwości, jakie daje tworzenie podsystemów i nie odnoszą się do kategorii życia. Dennett zgadza się z tym, że wyznając taka teorię musi zbudować pełne ja wyłącznie z modułów i podzespołów dostępnych w funkcjonalizmie. Zaznacza też, że budowany przez niego model jest modelem eklektycznym[63] (rys. 3).

Wyróżnionych zostało sześć modułów funkcjonalnych zajmujących się odpowiednio:

1. Prezentowanie informacji na wyjściu (PR), czego dokładniejsze wyjaśnienie Dennett pozostawia psycholingwistyce.
2. Introspekcją, przesyłaniem aktów mowy na wyjście i kontrolą stanu na wyjściu.
3. Przechowywaniem informacji w rodzaju pamięci krótkoterminowej (M), nazywanej też pamięcią buforową.
4. Analizą danych percepcyjnych, na którą składa się pamięć ikoniczna, równoległe procesy rozpoznawania cech oraz generowanie i testowanie hipotez.

[62] Por. D.C. Dennett, *Brainstorms*, s. 150.
[63] Tamże, s. 155.

5. Rozwiązywaniem problemów realizowanym również dzięki równoległym procesom rozpoznawania cech oraz generowaniu i testowaniu hipotez.
6. Wytwarzaniem snów, które podlegają takim samym interakcjom z systemem kontroli i pamięcią buforową jak dane z modułu analiz percepcyjnych.



A – sygnał z pamięci krótkoterminowej do modułu kontroli

M – pamięć krótkoterminowa

PR[64] – moduł komunikowania stanów wewnętrznych na wyjściu

Q – sygnał z modułu kontroli do pamięci krótkoterminowej

Rysunek 3.

[64] Z języka angielskiego: *Public Relations*.

Taki podział funkcji składających się na zdolności poznawcze jest, zdaniem Dennetta, podstawowy dla wyodrębnienia uwagi [*attention*]. Uwaga może być dla niego nieświadoma, co nie jest terminem sprzecznym. Występuje ona w stanach lunatycznych i w programach komputerowych. Operacje przeprowadzane na pewnych zbiorach danych mogą być przez teorię funkcjonalistyczną traktowane jako skupianie na nich uwagi. Dla wyjaśnienia fenomenu świadomości istotna jest jednak uwaga świadoma. Informacje pochodzące z doświadczenia percepcyjnego, z wyobrażeń towarzyszących rozwiązywaniu problemów oraz dotyczące planów, intencji, przekonań i celów biorących udział w procesie kontroli zostają wysłane do pamięci buforowej. Wraz z modułem kontroli ma ona zasadnicze znaczenie dla skupiania uwagi, ale jak zauważa Dennett, nie wnosi to jeszcze nic nowego dla znalezienia istotnej różnicy pomiędzy uwagą świadomą i nieświadomą.

W psychologii poznawczej toczy się debata pomiędzy tymi, którzy uważają, że tylko zdania [*propositions*] mogą być nośnikami informacji oraz tymi, którzy sądzą, że mogą nimi być zarówno zdania jak i obrazy [*images*]. Oddaje to, zdaniem Dennetta, wzajemną zależność pomiędzy filozofią, głównie epistemologią a psychologią empiryczną. Niemożliwość znalezienia odpowiedzi na padające w tej debacie pytania empiryczne dotyczące ludzkiego systemu nerwowego powoduje, że stawiane są pytania mniej empiryczne, związane tylko z mechanicznym i projektowym opisem funkcji poznawczych. Nie chodzi już o układ nerwowy człowieka, ale o jakikolwiek system zdolny spełnić te funkcje. Zdania lub obrazy jako nośniki informacji w systemach poznawczych mogą być rozważane na gruncie filozofii, psychologii, biologii lub cybernetyki. W modelu proponowanym przez Dennetta treść [*content*] zawarta w danym nośniku jest zależna od funkcji, jaką pełni ten nośnik w systemie jako całości.

Nie wszystkie funkcje spełniane w systemie poznawczym muszą być uświadamiane. Nie są uświadamiane analizy danych percepcyjnych biorące udział w rozpoznawaniu przedmiotów oraz wiele innych procesów. Zdaniem Dennetta mamy świadomy dostęp do ich rezultatów, lecz nie do nich samych. Przykładami tych rezultatów są wrażenia wewnętrzne [*internal impressions)* i wrażenia zmysłowe [*sensations*], które Dennett rozpatruje odwołując się do podziału dokonanego przez Hume'a[65]. Jest to jednak tylko luźne nawiązanie do tego podziału, gdyż Hume wyróżnił *impressions of sensation* oraz *impressions of reflexion*[66],o czym Dennett nie wspomina. Utożsamia on wprowadzone przez Hume'a wrażenia refleksyjne [*impressions of reflexion*] z wprowadzonymi przez siebie wrażeniami wewnętrznymi]*internal impressions*][67].

Spójność i stałość wrażeń zmysłowych uzasadnił Hume ciągłością istnienia ciał zewnętrznych. W odróżnieniu od wrażeń zmysłowych, które nie istnieją, kiedy ciała zewnętrzne nie są postrzegane, wrażenia wewnętrzne w koncepcji Hume'a istnieją, kiedy nie są postrzegane odpowiadające im ciała zewnętrzne. Błędem byłoby przypuszczać, że nie istnieją żadne, zapewniające ciągłość istnienia wrażeń wewnętrznych ciała zewnętrzne. Wyobrażenia mentalne [*mental images*] należy, zdaniem Dennetta, traktować analogicznie do wrażeń zmysłowych. Nasze doznanie zawiera na ogół więcej informacji, niż utworzone na jego podstawie doświadczenie zewnętrzne. Analiza danych zmysłowych w systemie percepcyjnym wykazuje, gdzie powinny one zostać zmodyfikowane dla lepszej kontroli zachowań, co nie jest dla podmiotu świadome, ani dostępne introspekcyjnie. Na podstawie

[65] Por. D.C. Dennett, *Brainstorms*, s.167.

[66] Por. D. Hume, *A Treatise of Human Nature*, s. 55.

[67] Termin *reflexion* tłumaczony może być jako: odbicie, odzwierciedlenie, namysł, zastanowienie, refleksja. Dennett przyjmuje najwidoczniej, że (jakiekolwiek tłumaczenie zostanie zastosowane) związane z *reflexion* wrażenie spełniane będzie wewnętrznie w umyśle.

działania systemu percepcyjnego, który potrafi dodawać do doświadczenia elementy, których nie ma w pierwotnym doznaniu można wnioskować o podobnej naturze wrażeń zmysłowych i wyobrażeń mentalnych.

3.4. Świadomość w sensie *awareness* i świadomość w sensie *consciousness*

Terminy: *consciousness i awareness* używane są zazwyczaj w języku angielskim jako synonimy. Dennett uważa, że powoduje to trudności w sformułowaniu poprawnej teorii świadomości. Postanawia on uściślić znaczenia tych terminów. Pierwszą rzeczą, którą należy zauważyć jest to, że mogą one być użyte intencjonalnie lub nieintencjonalnie. Użycie intencjonalne charakteryzuje się świadomością czegoś, a użycie nieintencjonalne ogólnie dotyczy bycia świadomym aktualnej sytuacji[68], na przykład takiej, że jest się świadomą formą życia. Dennett proponuje następnie unikanie stwierdzeń typu: „świadomy, że" [*consciouss that*] i tłumaczenie terminu „świadomy czegoś" [*consciouss of*] zawsze jako *aware of*. Dla spotykanego czasem nieintencjonalnego użycia terminu „świadomość", jako *awareness,* wprowadza tłumaczenie go na „czujność" [*alertness*], co można jego zdaniem podporządkować pod tę samą kategorię co nieintencjonalne użycie terminu „świadomość", zarówno w sensie *consciousness,* jak i *awareness*. Świadomość w sensie *consciousness* i świadomość jako *awareness* to dwa różne fenomeny, które Dennett analizuje osobno.

Świadomość pojęta jako *awareness* zależna jest od dwóch rodzajów cech. Jedną z nich jest nasza zależność od świadomości rzeczy w naszym otoczeniu dla poruszania się. Ze względu na reakcje zwierząt na środowisko uznajemy, że są one świadome [*aware*]. Nie interpretują jednak otaczających je rzeczy tak jak ludzie. Dogodniej jest przypisać świadomość ludziom, gdyż mogą oni opowiedzieć o tym, czego są świadomi. Stąd wiemy, że mają dostęp do treści swej świadomości. Nie zawsze jednak kontrola zachowania idzie u człowieka w parze z możliwością zakomunikowania kolejnych posunięć. Dennett wnioskuje stąd, że można je traktować rozłącznie.

Użycie terminu: „świadomość" [*awareness*] Dennett przekształca w jego formę właściwą dla postaw propozycjonalnych. Pisze on odtąd już tylko: „świadomy, że" [*aware that*], co zwykle wiąże się ze stwierdzaniem prawdziwości. Wyróżnia następnie dwa rodzaje świadomości [*is aware*]:

1. A jest świadomy-1, że *p* w czasie *t* wtedy i tylko wtedy, jeżeli *p* jest treścią stanu na wejściu centrum mowy A, w czasie *t*.
2. A jest świadomy-2, że *p* w czasie *t* wtedy i tylko wtedy, gdy *p* jest treścią pewnego zdarzenia wewnętrznego A w czasie *t,* które kieruje aktualnym zachowaniem.

Centrum mowy rozumie Dennett w funkcjonalnym lub logicznym, nie anatomicznym sensie. Zwierzęta nie posiadają go i są tylko świadome-2 [*aware-2*]. W ten sposób mogą również być świadome maszyny cybernetyczne. Jeżeli do maszyny takiej dołączone zostałoby centrum mowy, to byłaby ona świadoma-1 [*aware-1*]. Te dwa rodzaje świadomości wyczerpują całą przestrzeń możliwości tak, że nie ma potrzeby wprowadzania świadomości-3 [*awareness-3*]. W jednym z proponowanych systemów poznawczych, nazywanym maszyną percepcyjną, Dennett wprowadził fikcyjną linie podziału pomiędzy częściami funkcjonalnymi tej maszyny,

[68] Por. D.C. Dennett, *Content and Consciousness*, s. 115.

a więc częściami analizy aferentnej a centrum mowy. Tę linię świadomości [*awareness line*] rozumieć należy jako konstrukcję pojęciową nałożoną na system fizykalny. Treść sygnału, który nie przekroczył tej linii nie może zostać przez osobę zakomunikowana. Sygnał nie musi jednak przekraczać tej linii, by wywołać pewne odruchowe reakcje systemu nerwowego. Możemy być świadomi-1 tylko jednej rzeczy w danym czasie, a w tym samym czasie mózg kontroluje i jest świadomy-2 wielu innych funkcji. Świadomość-1 zwiększa skuteczność kontroli zachowania, gdyż w przypadku popełnienia błędu w postępowaniu kontrolowanym za pomocą świadomości-2 następuje swego rodzaju negatywne sprzężenie zwrotne, przywracające świadomość-1, która pozwala skorygować błąd. Zachowanie, które zostało poddane świadomej kontroli zostaje często zautomatyzowane i poddane kontroli wyspecjalizowanych systemów, które używają tylko świadomości-2.

Nieintencjonalne użycie terminu: „świadomość" jako *consciousness* również jest dwuznaczne. Dennett wyróżnia tu bycie świadomym ogólnej sytuacji oraz samą zdolność do bycia jej świadomym. Tylko o istotach zdolnych do bycia świadomymi można poprawnie powiedzieć, że są w jakimś momencie nieświadome. Świadomość w sensie samej tylko zdolności do bycia świadomym zależna jest od tego, na czym bycie świadomym polega. Wyjaśniając świadomość w tym bardziej podstawowym znaczeniu, zamiast tradycyjnego podziału na świadomość i nieświadomość, Dennett proponuje wprowadzenie stopni świadomości. Jeżeli uznać, że świadomym jest się w czasie snu, to będzie to świadomość-2 [*awareness-2*]. W tym samym sensie świadome są również zwierzęta. Jeżeli jednak za świadomość uznamy świadomość-1 [*awareness-1*], ani śpiący ani zwierzęta nie będą świadome.

Problem ten nieco inaczej wygląda w języku polskim, gdzie dla wyrażenia bycia świadomym mamy tylko jedno słowo. Zgodne z poglądami Dennetta jest przyjęcie, że dzięki temu unikamy problemów teorii świadomości związanych z zamieszaniem terminologicznym. Zachować należy podział na świadomość-1 i świadomość-2 oraz pamiętać, że wprowadzenie linii świadomości jest tylko pojęciowym uproszczeniem. W rzeczywistości świadomość doświadczana jest w postaci stopni mniejszego lub większego uświadomienia.

3.5. Wyobrażenia i jaźń

Zgodnie z przedstawionym przez Dennetta poglądem na świadomość, nie jesteśmy świadomi, w żadnym sensie tego słowa, obrazów mentalnych[69]. Introspekcja pokazuje jednak, że świadomość wypełniona jest różnymi przedmiotami. Dennett przedstawia argumenty na to, że pojawianie się tych przedmiotów wytłumaczyć można środkami czysto fizykalnej teorii umysłu. Robi to na przykładzie percepcji wzrokowej zakładając, że interesujące go rezultaty stosują się również do innych zmysłów.

Wyobrażenia [*images*] są subiektywne i intencjonalne, nie mają wymiarów przestrzennych. Wyobrażenie jest zawsze reprezentacją [*representation*] czegoś. Ma ono przynajmniej jedną jakość [*quality*], kształt, kolor lub formę wspólną z tym, co jest reprezentowane. Nie spełnia ono wyłącznie roli symbolicznej lub funkcjonalnej, lecz jest do reprezentowanej rzeczy podobne. Dennett rozumie wyobrażenia w ten sposób, że ostatnim z wyobrażeń, najbliższym światu zewnętrznemu jest pobudzenie siatkówki oka. Na jej powierzchni zaczyna się proces

[69] Tamże, s. 132.

analizy aferentnej i biegnie dalej poprzez nerw optyczny. Wzór obrazu z siatkówki zostaje zamieniony na odpowiednie informacje. Jeżeli miałyby one zostać ponownie zamienione na ten określony wzór w mózgu, to proces percepcyjny musiałby zostać powtórzony od początku, począwszy od takiego samego obrazu, jak wzór z siatkówki oka. Prowadziłoby to do regresu do nieskończoności. Dennett pisze o reprezentacjach neuronalnych, które nie są obrazami reprezentowanych rzeczy, lecz wzorami pewnych połączeń neuronalnych. Na takim subosobowym poziomie wyjaśniania procesu percepcyjnego nie ma miejsca na wyobrażenia.

Opis wyobrażeń na poziomie osobowym Dennett rozpoczyna od przytoczenia poglądu J. M. Shortera, który w *Imagination* uznał wyobrażenia za bliższe opisom słownym niż obrazom[70]. W wyobrażeniu pewne jego elementy mogą zostać całkowicie pominięte, nie jak ma to miejsce w przypadku obrazu, tylko niewyraźnie zarysowane. Za opisem słownym przemawia również percepcja zamieszczonego przez Dennetta rysunku, na podstawie którego rozważał percepcję Wittgenstein[71]. Przedstawione na rysunku zwierzę może wyglądać jak kaczka lub jak królik. W przypadku tym nie ulega zmianie wzór na siatkówce, ani na papierze. Jego świadomość-1 [*awareness-1*] zależna jest od treści sygnałów przekraczających linię świadomości.

Dennett zwraca również uwagę na związek wyobrażania z widzeniem. Przemawia za tym fakt, że wyobrazić można sobie tylko to, co możliwe byłoby do zobaczenia gdyby istniało. Wyobrażanie podobne jest raczej do widzenia niż do malowania obrazu, gdyż nie można w wyobrażeniu, ani w percepcji wzrokowej odróżnić figury, która ma tysiąc boków od figury, która ma ich dziewięćset dziewięćdziesiąt dziewięć. Można to zrobić licząc boki, a wtedy sygnał musi dostać się do centrum mowy. Dla potwierdzenia udziału centrum mowy i świadomości-1 w procesie widzenia i wyobrażania Dennett podaje przykład koncentrowania wzroku tylko na małej, centralnej części pola widzenia. Jego reszta pozostaje tylko półświadoma [*semi-aware*]. Możliwość rozpoznania i nazwania otaczających centrum i widzianych wcześniej niewyraźnie przedmiotów, pojawia się dopiero w momencie skupienia na nich uwagi.

Jaźń [*self*] uważana była za niefizykalną duszę, rodzaj ducha w maszynie lub bezprzedmiotowe wyobrażenie. Dennett nie kwestionuje istnienia jaźni i chce pokazać jej wytworzenia na drodze ewolucji. Granica jaką wytworzyły proste organizmy pomiędzy sobą a światem zewnętrznym jest również granicą oddzielającą od tego świata jaźń. Dennett ma tu na myśli prostą jaźń biologiczną, którą nazywa abstrakcyjną zasadą organizacji[72]. Zwierzęta chronią granic swych terytoriów, czego rezultatem jest rozszerzony fenotyp [*extended phenotype*]. R. Dawkins określa go jako podstawowe wyposażenie biologiczne indywiduów. Definicja rozszerzonego fenotypu rozciąga granice indywiduum poza jego cielesność. Włączone zostają konstrukcje i wyposażenia zewnętrzne, organizmy żyjące z indywiduum w symbiozie i inne indywidua tego samego gatunku. Najbardziej zadziwiającą konstrukcją służącą rozszerzeniu i umocnieniu granic jest ludzka jaźń. Dennett uważa ją za taki sam wytwór biologiczny jak pajęczyna pająka, skorupa żółwia, czy muszla ślimaka.

Organizacja jaźni ludzkiej sprawia na obserwatorze wrażenie, że obserwowana osoba kierowana jest jakąś nadrzędną duszą. Dennett chce wykazać błędność tego wrażenia przyrównując organizację jaźni do kolonii termitów. Ona również wywiera wrażenie posiadania duszy będąc tylko złożeniem automatów. Od takich prostych organizmów ludzie

[70] Tamże, s. 135.
[71] Tamże, s. 137.
[72] Por. D.C. Dennett, *Consciocness Explained*, s. 414.

różnią się prezentowaniem siebie innym i sobie samym, co odbywa się za pomocą słów – memów kultury. Samoobrona i samokontrola spełniana jest tutaj poprzez opowieści. Za wyjątkiem ludzi zajmujących się tym profesjonalnie, zwykle nie rozważamy, co opowiedzieć, ani jak to zrobić. Zdaniem Dennetta to nasze opowieści określają jacy jesteśmy, a nie my je określamy. Podobnie jak w fizyce uproszczeniem jest wyznaczenie centrum grawitacji przedmiotu, uproszczeniem jest wskazanie centrum narracyjnej grawitacji [*center of narrative gravity*][73].

Dennett podaje przykład rozszczepienia osobowości, w którym na jedno ciało przypada więcej niż jedna jaźń. Prawie we wszystkich badanych przypadkach rozszczepienie nastąpiło na skutek bolesnych doświadczeń z wczesnego dzieciństwa. Badane osoby zdolne były wykształcić dodatkową jaźń lub jaźnie, których wspomniane doświadczenia nie dotyczyły. Z drugiej strony Dennett wspomina również o przypadku bliźniąt, które są przykładem istnienia jednej jaźni dla dwóch ciał. Opisane osoby potrafiły zachowywać się bardzo podobnie, jedna potrafiła kończyć zaczęte przez druga zdanie. Jest to zdaniem Dennetta argument na istnienie wielu trudnych do rozpoznania sposobów komunikacji i koordynacji zachowań. Bliźnięta, które urodziły się z tak samo wyposażonymi mózgami i doznały w ciągu swego życia mniej więcej tego samego mogą z takich sposobów korzystać. Jaźń tak jak świadomość, choć sprawia wrażenie ciągłości zawiera przerwy i wznowienia. Może też dochodzić do sytuacji, kiedy części jaźni tracą zdolność do komunikacji. Przykładem podanym przez Dennetta jest tutaj przecięcie połączenia pomiędzy półkulami mózgowymi.

Spostrzeżenia Dennetta są ciekawe i zarysowują kierunki przyszłych badań. Problemy świadomości i jaźni wciąż pozostają nierozwiązane, choć proponowane przez Dennetta podejście umieszcza je w zakresie ludzkich możliwości poznawczych. Problemem jest tylko stopień komplikacji i interdyscyplinarność zagadnień. Można zastanawiać się nad ich aspektem etycznym, gdyż celem jest tutaj stworzenie sztucznego człowieka. Z drugiej strony badania takie mogą służyć doskonaleniu ludzkich możliwości, kompensowaniu ułomności i kalectwa. Nie trzeba więc zastanawiać się nad ich aktualnością.

4. Język i sztuczna inteligencja

4.1. Język mózgu i język umysłu

Odrzucając ideę teatru kartezjańskiego Dennett musi pokazać jak zdania i wypowiedzi mające określone znaczenie dla podmiotu powstają bez udziału wewnętrznego obserwatora. Wprowadzenie takiego kolejnego podmiotu powoduje, zdaniem Dennetta, problem regresu do nieskończoności. Pokazywanie, w jaki sposób można go pominąć rozpoczyna on od uwagi, że wszystkie teorie lingwistyczne skupiają się na odbiorze wypowiedzi językowych, nie mówiąc zbyt wiele o ich wytwarzaniu. Fala dźwiękowa lub wzrokowe doznanie znaków językowych jest tylko materiałem wejściowym, na którym pracują systemy: percepcji i rozumienia. Dennetta interesuje również materiał wyjściowy systemu rozumienia mowy [*speech comprehension*]. Zastanawia się on, czy materiał ten składa się ze zdań języka mentalnego, (w

[73] Tamże, s. 418.

terminologii J. Fodora – *Mentalese*), obrazów w głowie, pewnych zbiorów struktur głębokich, czy też jakichś niewyobrażalnych rzeczy[74].

Przytacza badania duńskiego lingwisty Pima Leveta, który zajmuje się procesem wytwarzania wypowiedzi[75]. Levet twierdzi, że mowa nie powstaje w procesie planowania i wypowiadania jednego słowa w danym czasie. System wytwarzania mowy posiada zdolność przewidywania, czego przykładem może być dopasowywanie akcentu do słów, które mają dopiero pojawić się na wyjściu tego systemu oraz częstsze występowanie przejęzyczeń mających formę rzeczywiście funkcjonujących słów od takich, które są tylko do słów podobne i nic nie znaczą. Wprowadzony przez Leveta generator pojęć [*conceptualizer*], choć z pewnego punktu widzenia może przypominać odrzucaną przez Dennetta jednostkę centralną, stwarza możliwość wyjaśnienia sposobu nadawania słowom znaczenia bez użycia takiej jednostki. Można to pokazać poprzez błędy popełniane przez system, gdyż uwidaczniają one współpracę różnych jego modułów. Generator pojęć podaje nie zwerbalizowaną jeszcze wiadomość do modułu nadającego jej określoną formę gramatyczną i fonetyczną. Częstsze występowanie błędów przybierających postać rzeczywiście funkcjonujących słów lub wyrażeń tłumaczone jest wpływem informacji zawartej w tym module. Nie zwerbalizowana wiadomość musi docierać tam w formie mentalnego lub mózgowego języka zrozumiałego dla podsystemów generowania pojęć, lecz różnego od języka pojawiającego się na wyjściu systemu. Język myśli [*language of thought*] zawiera szczegółowe informacje potrzebne dla wytworzenia wypowiedzi w języku naturalnym, takim jak język polski lub angielski. Dennett rozważa możliwość nadania językowi myśli postaci kodu binarnego, który zamieniany jest następnie na cyfry lub litery.

Zamiast wprowadzonego przez Leveta generatora pojęć, który jest homunculusem odpowiedzialnym za wiele funkcji i posiadającym odpowiednio wiele informacji, wprowadzić można całe pandemonium złożone z demonów, z których każdy specjalizuje się w spełnianiu jednej, przypisanej mu funkcji. Określony demon odpowiedzialny będzie za zmianę języka myśli w język wypowiedzi, a inny za przypisanie tej wypowiedzi treści. Dennett zauważa, że problemem okazuje się wtedy ustalenie sposobu komunikacji pomiędzy poszczególnymi demonami. Można założyć minimum współpracy i wtedy demony językowe automatycznie nadają nie zwerbalizowanej wiadomości określoną formę, bez rozważania alternatywnych możliwości. Drugim skrajnym rozwiązaniem jest przyjęcie różnorodnych oddziaływań pomiędzy możliwymi formami leksykalnymi, dźwiękami, znaczeniami i formami gramatycznymi. W takim ujęciu intencja towarzysząca procesowi komunikowania się jest zarówno jego przyczyną jak i skutkiem. Wypowiedź w żadnym jej momencie nie jest określona do końca. Zależność wypowiedzi od szerokiego kontekstu związanego z funkcjami różnych demonów może spowodować dołączenie do niej różnych treści i nadanie różnych form gramatycznych. Proces taki wymaga więc systemu kontroli. Odpowiedź na pytanie, ku której ze skrajności bliższe jest rzeczywiste generowanie wypowiedzi językowych, nie została jeszcze poznana. Dennett skłania się jednak ku pandemonium, w którym przebiega wiele procesów równoległych, w żadnym momencie nie określających wypowiedzi ostatecznie.

Przejęzyczenia freudowskie są zdaniem Dennetta przykładem nieuświadomionego działania maszyny von neumannowskiej, która testuje wszystkie słowa i zwroty podobne do danego przykładu, a następnie wybiera jeden najciekawszy lub najbardziej trafny i nadaje mu formę świadomej wypowiedzi. Ma to świadczyć o przebiegu wielu procesów równoległych, istotnych dla twórczego użycia języka. Podmiot formułujący wypowiedzi nie jest ich od razu

[74] Por. D.C. Dennett, *Consciousness Explained*, s. 231.
[75] Tamże, s. 232.

świadom, a więc nie wie, co myśli, zanim nie znajdzie odpowiednich dla ich wyrażenia słów. Nie zwerbalizowana informacja zostaje przez podmiot poznana dzięki nadaniu jej określonej formy językowej. Samo wybieranie odpowiednich słów, które odbywa się w szerokim kontekście związków znaczeniowych i różnego rodzaju podobieństw, ma znaczenie dla treści ostatecznie uświadomionej wypowiedzi. Zaznacza, że nie odrzuca całkowicie możliwości sprawdzenia się modelu generatora językowego jako jednego podsystemu. Podaje jednak argumenty, które jego zdaniem przemawiają za istnieniem wielu współdziałających demonów językowych. Problemem wydaje się niemożliwość empirycznego sprawdzenia działania takiego modelu, gdyż wymagałaby ona eksperymentów na mózgu żywego człowieka. Dennettowi muszą wystarczyć argumenty oparte na obserwacji zaburzeń zachowań językowych, które powstają przy uszkodzeniach układu nerwowego. Dopuszczają one jednak możliwość różnych interpretacji. Nowych rozwiązań dostarczyć mogą techniki monitorowania pracy mózgu, podczas różnych aktywności podmiotu.

Stary pomysł na język myśli powinien zostać rozwinięty do postaci szczegółowej hipotezy dotyczącej procesów poznawczych. Żadna taka testowalna hipoteza nie została jeszcze zaproponowana. Dennett rozważa hipotetycznie sprowadzenie przekonań do postaci zdań wyrażających postawy propozycjonalne. Zakłada następnie algorytm pozwalający sprowadzić je do stanów komputacyjnych i umiejscowienia ich w *hardware* organicznego mózgu. Wtedy okazałoby się, że przekonania i inne postawy propozycjonalne są, jak to określił Fodor[76]: „(...) w jakiejś *komputacyjnej* relacji do pewnej wewnętrznej reprezentacji[77].” Może jednak również okazać się, że dla działania systemu percepcyjnego i motorycznego nie jest potrzebny język myśli. Wyższe funkcje poznawcze, takie jak: planowanie, rozwiązywanie problemów, czy ustalanie przekonań, różnią się od percepcji i działania tym, że wymagają myślenia. Powinny więc wymagać również pewnego języka myśli, formy mózgowego zapisu treści. Alternatywą może okazać się tutaj koneksjonizm.

Proponowane przez koneksjonizm sieci połączeń złożone są ze względnie prostych elementów. Dennett podaje kilka różnic pomiędzy takimi sieciami, a tradycyjnymi modelami nauk poznawczych[78]:

1. „rozdzielona” pamięć i przetwarzanie, w których części odgrywają różne, niezwykle wieloznaczne role, i w których ustalenie znaczenia pojawia się tylko „globalnie”, (w skrócie, nie ma zdań [*propositions*] w określonych miejscach [*addresses*] w pamięci);
2. żadnej centralnej kontroli, a raczej częściowo anarchiczny system poniekąd współzawodniczących elementów;
3. żadnego złożonego przekazywania wiadomości pomiędzy częściami składowymi;
4. pewne zaufanie do statystycznych własności całości w osiąganiu rezultatów;
5. stosunkowo pozbawione umysłu i niewydajne dokonywanie i niedokonywanie wielu częściowych połączeń lub rozwiązań, aż system po chwili osiągnie stan ustalony – niekoniecznie zgodny z przewidywaną „właściwą” odpowiedzią.

Modele koneksjonistyczne są komputacyjne w tym sensie, że realizuje się je przy użyciu podzespołów komputerowych. Zachowanie się każdej części modelu jest dokładnie określona funkcją zachowania się pewnych innych jego części. Modele te różnią się jednak od modeli tradycyjnie pojętych nauk poznawczych tym, że bliższe są naukom neuronalnym niż psychologii. Nie są komputowane zdania [*propositions*] jakiegoś rachunku predykatów, ani

[76] J. Fodor, *The Language of Thought*, s. 198.
[77] Por. D.C. Dennett, *Intentional Stance*, s. 228.
[78] Tamże, s. 229-230.

opisy formalne pewnych transformacji gramatycznych, ale na przykład, nowe wartości progowe elementów, które same przez się nie mają żadnego określonego znaczenia, ani semantyki świata zewnętrznego. Na tym poziomie opisu semantyka odnosi się tylko do zdarzeń, procesów, stanów i miejsc w mózgu, w samym tylko systemie komputacyjnym.

Na wyższym poziomie opisu przypisujemy własności semantyczne świata zewnętrznego aktywności sieci neuronalnej. Na tym poziomie jednak związki i interakcje pomiędzy elementami semantycznymi nie są komputacyjne, ale jak pisze Dennett, statystyczne, emergentne i holistyczne. Zachowanie się takiej maszyny wirtualnej nie może zostać sformalizowane przy użyciu słownictwa poziomu psychologicznego, które jest komputacją algorytmu wyższego rzędu. Poziom algorytmiczny nie może tutaj, zdaniem Dennetta, przypominać zwykłego języka maszyny. Nie zakłada się bezpośredniego tłumaczenia wyższego poziomu posiadającego semantykę świata zewnętrznego, na niższy poziom algorytmiczny. W modelach koneksjonistycznych *hardware* dodaje coś, co na wyższym poziomie może zostać przypisane tylko statystycznie.

Proponując to Dennett próbuje określić miejsce swego stanowiska, pomiędzy innymi proponowanymi koncepcjami. Rozwiązania proponowane przez Fodora, Sticha i Churchlandów podaje jako przykład eliminatywizmu, z którym się nie identyfikuje. Fodorowi przypisuje podkreślenie bliskiego związku języka myśli z realizmem uwzględniającym postawy propozycjonalne. Zastąpienie psychologii ludowej [*folk psychology*] teorią syntaktyczną uznaje za właściwe Stichowi, a u Churchlandów zwraca uwagę na przyjętą metodologię. Odróżnia od tych wymienianych dla porównania stanowisk swoje własne. Koncentruje się w nim na szukaniu powiązań i oddziaływań pomiędzy poziomami nastawień: intencjonalnego, projektowego i fizykalnego.

4.2. Treść, rozumienie i informacja

Uzyskanie wystarczającej wiedzy na temat pracy mózgu i złamanie kodu cerebralnego jest przez niektórych filozofów uznawane za zadanie możliwe do zrealizowania. Dennett przytacza argument Anscombe, która jest przeciwna tej możliwości[79]. W *Intention* przedstawiła ona pogląd, zgodnie z którym informacje na temat przekonań lub intencji podmiotu można uzyskać tylko na podstawie informacji dotyczących działań i doświadczeń tego podmiotu. Nieprzydatne są do tego celu dane neurologiczne. Dennett nie zgadza się z tym argumentem pisząc, że każda poddana rozważaniu biografia, rozumiana jako zapis działań i doświadczeń podmiotu, jest całkowicie spójna w perspektywie więcej niż jednego zbioru intencji i przekonań. Zastanawia się następnie nad możliwością odkrycia systemu, którego mózg używa do przechowywania przekonań i intencji. Znajomość takiego systemu wraz ze znajomością zachowań podmiotu pozwoliłaby ustalić jego przekonania i intencje. Charakterystyczne dla takiego systemu stany mózgu byłyby przypisane określonym stanom umysłu. Hipoteza zapisu mózgowego [*brain-writing*][80] jest zdaniem Dennetta obiecująca, ale należy rozważyć implikowane przez nią problemy.

W modelu zaproponowanym przez Gilberta Harmana[81] język wewnętrzny zostaje utożsamiony z systemem reprezentacji. Zachowanie ludzkie zależne jest od tych

[79] Por. D.C. Dennett, *Brainstorms*, s. 39.
[80] Tamże, s. 40.
[81] Opiera się Dennett na: G. Harman, *Language Learning*, w: "Nous", IV/1970, s. 35.

wewnętrznych reprezentacji, które składają się na wewnętrzny język osoby. Dennett zastanawia się, czy każdy system reprezentacji musi mieć charakter języka i jak miałby on zostać zrealizowany w mózgu. Istotne jest również określenie roli przyczynowej reprezentacji dla zachowań zewnętrznych podmiotu. Należy brać pod uwagę możliwość istnienia tak nieprzeniknionej natury mózgowej realizacji systemu reprezentacji, że neurologia ani żadna inna nauka nie będzie zdolna ich wykryć. Dennett podaje jednak sześć warunków, które jego zdaniem powinny określić, na czym polega hipoteza zapisu mózgowego[82]:

1. Po pierwsze, system reprezentacji musi posiadać swoją gramatykę generatywną. Oznacza to, że jeżeli rozumie się ten system i zna jego skończone słownictwo, to wygenerować można reprezentacje, które będą zdaniami zapisu mózgowego. Nie musi to jednak być tylko jedna gramatyka generatywna. System może być wielojęzyczny i wtedy pewne jego funkcjonalne elementy służyć będą jako interpretatory.
2. Po drugie, syntaktyczne różnice i podobieństwa muszą znajdować swoje odbicia w fizykalnych różnicach i podobieństwach mózgu. Egzemplarze danego typu syntaktycznego muszą więc być fizycznie odróżnialne od egzemplarzy innego typu. Nie oznacza to jednak, że wszystkie egzemplarze danego typu muszą być fizycznie podobne. Wystarcza skończona ilość fizycznych rodzajów egzemplarzy każdego typu, jak dla pisanych i mówionych egzemplarzy słów.
3. Egzemplarze muszą być fizykalnie znaczące, choć wiadomo, że obecnie stosowane metody nie wystarczają dla ich rozpoznawania.
4. Zasób reprezentacji musi spełniać warunek spójności biograficznej podany przez Anscombe. Zdania podane w kodzie cerebralnym muszą pasować do przekonań i pragnień.
5. Musi istnieć pewien mechanizm pokazujący jak system fizykalny, w którym zrealizowany jest zapis mózgowy został funkcjonalnie powiązany z przyczynami aktywności cielesnej.
6. Zasób przekonań musi być spójny, choć dopuszczalne są pewne niewielkie niezgodności.

System spełniający te warunki zajmuje określona ilość miejsca w mózgu. Dennett przytacza oszacowanie, którego dokonał Minsky. Zgodnie z nim mózg z miliardami neuronów, zawiera wystarczająco dużo miejsca dla zaistnienia inteligencji. Podmiot posiadający ten mózg może jednak w nieskończoność mnożyć swe przekonania, generować różne ich odmiany i połączenia. Na tej podstawie Dennett wnioskuje, że nie są one przechowywane w mózgu w gotowej postaci, lecz istnieje tam generujący je mechanizm. Mechanizm ten nazywany przez Dennetta mechanizmem ekstrapolacyjno-dedukcyjnym [*extrapolator-deducer mechanism*], nie może polegać na informacji metajęzykowej, gdyż prowadziłoby to do regresu do nieskończoności. Potrzebne byłyby kolejne informacje, do których odwołają się właściwe wynikania i wnioskowania. Dennett podaje również przykłady świadczące o tym, że przechowywanie informacji w mózgu nie musi być związane z dołączaniem do niej przekonań. Formułowanie zdania wyrażającego przekonanie może je upraszczać, a zdania zawierające wieloznaczne zwroty i słowa są wyrażeniami poprawnymi dla dwóch różnych przekonań. Ich podmiot jednak zawsze wie, o które z przekonań chodzi, co zdaniem Dennetta świadczy o istnieniu wewnętrznego języka zapisu, różnego od języka naturalnego. Mózg jest organem, który reprezentuje, lecz jego reprezentacje nie mają postaci zdań. Dennett zastanawia się, czy mogą one mieć postać możliwego do nauczenia się [*learnable*] systemu, we wprowadzonym przez Davidsona sensie takiego systemu.

---

[82] Por. D.C. Dennett, *Brainstorms*, s. 42-43.

Zadaniem koncepcji centralistycznej, której przedstawicielem jest Dennett, jest wyjaśnienie sposobu przypisywania zdarzeniom znaczenia. Zostaje to zrealizowane poprzez nadanie treści określonym wypowiedziom słownym. Dennett rozważa powiązanie treści, szeroko pojętej jako zdarzenia w organach zmysłów, z treścią dotyczącą zdarzeń i przedmiotów ze świata zewnętrznego. Powiązanie to może być spełniane dzięki pamięci wcześniejszych doświadczeń i ich skutków. Zwierzęta z pewnością nie różnicują świata zewnętrznego w taki sam sposób, w jaki robią to ludzie. Nie podporządkowują one otaczających je przedmiotów pod pojęcia, które można by porównać do pojęć ze słownika jakiegokolwiek języka naturalnego. Niektórzy z centralistów proponują sformułowanie sztucznych języków, którymi posługują się zwierzęta. Zdaniem Dennetta posunięcie takie grozi zagubieniem sensu centralizmu, który ma powiązać pewne wyjaśnienia i opisy intencjonalne z pewnymi wyjaśnieniami i opisami ekstensjonalnymi. Będzie to niemożliwe, jeżeli niesłusznie przypiszemy danemu podmiotowi posługiwanie się jakimś językiem.

Ciekawe jak proponujący heterofenomenologię Dennett rozwiązałby problem znaczenia zdarzeń zewnętrznych, dla podmiotów będących przedstawicielami gatunków innych niż gatunek ludzki, z jego specyficznym umysłem. Ciekawe, czy mogą istnieć umysły tak różne od tych, które znamy, że umysły nasze nie posiadają i nie będą mogły posiadać narzędzi stosownych do ich poznania i zrozumienia? Heterofenomenologia może być skuteczna dla umysłów potencjalnie mogących przypisywać te same typy znaczeń do danych zdarzeń. Jaki jednak należy wybrać typ znaczenia dla umysłu różnego od umysłu człowieka? Umysłu korzystającego z innego rodzaju percepcji nawet, jeżeli korzysta on z tej samej logiki? Rozwiązaniem może być tu zastosowanie hipotezy ewolucyjnej, która traktuje wszystkie umysły jako różne wersje i stadia rozwoju tego samego umysłu. Dzięki temu można, mniej lub bardziej dokładnie, wyobrazić sobie pewne stany wewnętrzne spełniające role znaczeń. Ciągle jednak pozostaje problem umysłów zbyt odmiennych od ludzkich, by można było wyobrazić sobie ich stany wewnętrzne.

Możliwe jest, że w mózgu rozwinięte zostały metody przechowywania i przekazywania informacji oparte o zdarzenia lub struktury posiadające określoną i istotna dla tych metod syntaktykę. Słowami języka mózgowego będą wtedy wzory sygnałów neuronalnych. Do odczytania złożonego z takich znaków zapisu mózgowego potrzebny jest odpowiedni mechanizm. Musi on być systemem przetwarzania informacji. System taki jest, zdaniem Dennetta, złożony ze współpracujących ze sobą podsystemów, które przechowują i przekazują informacje w formie zapisu mózgowego.

Osoba rozważana jako system intencjonalny ma przekonania, zna fakty i przetwarza informacje. Ich zbiór jest zawsze zbiorem otwartym i może zostać powiększony o nowe fakty oraz informacje otrzymane w wyniku procesów ich przetwarzania. Względne jest, zdaniem Dennetta, odróżnienie wiedzy od przekonań. Przekonania natomiast należy uszeregować pod względem stopnia ich pewności. Podobnie należy postąpić z wiedzą, gdyż problemem jest ustalenie, jaką treść dany podmiot przypisuje wypowiadanym zdaniom. Dla określenia stopnia wiedzy przydatne może być rozróżnienie zaproponowane przez Ryle'a[83], na wiedzę „że" i wiedzę „jak", którą sprawdzić można w działaniu.

Związanie zdania z samym tylko stanem rzeczy we wszechświecie, pomija udział podmiotu w procesie przetwarzania informacji. Podmiot ten wypowiada lub odbiera i interpretuje zdania opisujące ten stan rzeczy. Uwzględnić należy również cechy funkcjonalne systemu

[83] Por. D.C. Dennett, *Content and Consciousness*, s. 184.

intencjonalnego, jakim jest ten podmiot. W teorii informacji istotny jest na przykład pomiar niezawodności kanałów jej przesyłania lub systemów. Ilość informacji określana jest na ogół przez zmniejszenie stopnia niepewności u jej odbiorcy. Zależy więc ona od informacji wstępnie posiadanej przez odbiorcę i będzie tym większa, im mniej informacji posiadał odbiorca w stanie początkowym.

4.3. Filozoficzne problemy sztucznej inteligencji

Rozwiązania proponowane przez badaczy sztucznej inteligencji mogą być przydatne dla rozstrzygnięcia pewnych kwestii występujących w filozofii umysłu. Dennett, choć nie jest w tej dziedzinie specjalistą, postanawia podać przegląd zagadnień z zakresu sztucznej inteligencji, które mogą rzucić nowe światło na problemy związane z filozofią umysłu. Pisze on, że filozofia umysłu i sztuczna inteligencja są bardzo podobne z punktu widzenia psychologów eksperymentalnych. Wykorzystują one analizę pojęciową i rozumowania aprioryczne, a jedyną pomiędzy nimi różnicą jest to, że sztuczna inteligencja sprawdza działanie projektu umysłu.

Podejmowane przez psychologię próby wyjaśnienia ludzkiego poznania, uczenia się, czy percepcji, zdaniem Dennetta zawsze prowadzą do połączenia psychologii z filozofią. Może to następować na dwa sposoby. Jednym z nich jest strategia budowania z pewnych podstawowych modułów, które nazwać można atomami psychologicznymi, coraz bardziej złożonych kompleksów. Drugi ze sposobów to rozkładanie wyższych poziomów organizacji psychologicznej na prostsze elementy. Podsystemy lub wyodrębnione procesy mają być ostatecznie sprowadzone do poziomu, którym zajmuje się biologia. Obie praktyki badawcze powinny być prowadzone równolegle, ale zdaniem Dennetta strategia prowadząca od wyższych poziomów do niższych nie jest zbyt obiecująca. Najlepiej rozwiniętymi programami badawczymi prowadzącymi od poziomów niższych do wyższych są: behawioryzm typu bodziec – reakcja i tak zwana przez Dennetta, „psychologia fizjologiczna sygnałów neuronalnych”[84]. Problemem, z którym spotyka się behawioryzm jest niejasność, co do statusu bodźca jako atomu psychologicznego. Dla „psychologii fizjologicznej sygnałów neuronalnych”, choć przebiegi impulsów neuronalnych poprzez synapsy są doskonałymi atomami psychologicznymi, to jest ich zbyt wiele, by poddać je efektywnemu badaniu. W wyniku wystąpienia tych problemów psychologowie starają się zadawać pytania wymagające mniej empirycznych odpowiedzi. Poszukują oni raczej rozwiązań technicznych niż odkryć. Jednak im więcej empirycznych ograniczeń nakładają oni na rozważany system, tym bardziej realne psychologicznie jest rozwiązanie.

Badania sztucznej inteligencji skupiają się na różnych poziomach empirycznych trudności. Może to być poziom danych empirycznych dotyczących ludzkich zachowań lub bardziej abstrakcyjny poziom możliwości nabycia wiedzy. Zawsze jednak proponowany jest przeciwny rozwiązaniom witalistycznym i dualistycznym, model mechanistyczny. Mechanicyzm ten nie jest dodatkowym wymogiem, lecz warunkiem stawianym przez psychologię. Sztuczna inteligencja jest więc takim samym abstrakcyjnym badaniem zasad psychologii, jak epistemologia lub filozofia umysłu. Od filozofii różni ją projektowanie odpowiednich dla spełniania określonych funkcji systemów, czym zajmuje się również

[84] Por. D.C. Dennett, *Brainstorms*, s. 110.

psychologia poznawcza. Filozofowie, jak pisze Dennett, zajmują się bardziej ogólnymi problemami.

Projektowanie systemów spotyka się z pewnymi zagrożeniami. Zaprojektowany może zostać system, którego podsystemom przypisano spełnianie niewłaściwych funkcji. Podsystemy te mogą zostać pomyślane jako posiadające zbyt wielkie możliwości przetwarzania informacji lub jako posiadające możliwości większe niż system, którego są częściami. W zbyt wielkie możliwości wyposażony został, zdaniem Dennetta, podsystem freudowskiego ego. Nastąpić może również pomylenie warunków wymaganych dla pewnego szczególnego rozwiązania z warunkami wymagań ogólnych. Na przykład warunki wymagane dla spełniania danych funkcji w umysłach ludzkich nie muszą być takie same, jak warunki wymagane dla spełniania tych funkcji w umysłach niższych gatunków zwierząt. Kolejnym błędem jest nałożenie projektu jednego z podsystemów na całą istotę poznawczą. Dennett podaje tutaj za przykład wczesny, napędzany syntaktycznie [*syntaks-driven*][85] system Chomsky'ego. Trudno wyjaśnić, w jaki sposób system ten wchodził w interakcje ze składowymi semantycznymi, by wytworzyć akty mowy. Można uznać, że jego funkcje rozciągnięte zostały również na te składowe. Ostatnim z podanych zagrożeń jest ograniczenie działania systemu tylko do jednej jego części, bez możliwości powiększenia systemu. Metoda zapamiętywania niewielkiej ilości informacji przez mechaniczne powtarzanie, nie może być zastosowana w przypadku większych całości treściowych.

Dennett proponuje, by wstępnie przyjąć, że rozkład funkcji w projekcie komputera jest izomorficzny z rozkładem funkcji w mózgu. Części funkcjonalne układu nerwowego mogą więc być zrealizowane w podzespołach elektronicznych. Komputer opisywany na głębokim poziomie jego konstrukcji jest urządzeniem cyfrowym, lecz rozważany na innych poziomach może symulować urządzenie analogowe. Na ogół mózg rozważany jest jako taki właśnie typ urządzenia. Nie można jednak mówić o psychologii modeli sztucznej inteligencji, ponieważ byłoby to przypisaniem im zbyt wielu funkcji. Można natomiast, zdaniem Dennetta uznać, że sztuczna inteligencja obala filozoficzno – psychologiczny argument mówiący, że jedyna psychologia, która mogłaby wyjaśnić złożoność aktywności ludzkiej musi używać wewnętrznych reprezentacji. Uważali tak prawie wszyscy, za wyjątkiem radykalnych behawiorystów. Dla Hume'a reprezentacjami były idee i wrażenia. Później psychologowie poznawczy zaczęli mówić o hipotezach, mapach, schematach, twierdzeniach, sygnałach neuronalnych.

Nic nie jest reprezentacją wewnętrznie, jest nią zawsze dla kogoś. Każda reprezentacja lub system reprezentacji wymaga przynajmniej jednego użytkownika lub interpretatora. Musi on być zdolnym do robienia porównań, mieć przekonania i cele, a więc być, jak pisze Dennett, czymś w rodzaju homunculusa. Psychologia bez homunculusów jest niemożliwa, a psychologia z homunculusami prowadzi do regresu do nieskończoności.

Jest to, zdaniem Dennetta, stosunkowo wyabstrahowana wersja pewnej grupy problemów. Dennett nazywa ją problemem Hume'a, ponieważ jego wrażenia i idee uznać można za wewnętrzne reprezentacje. Hume próbował uniknąć wprowadzenia pojęcia jaźni wewnętrznej [*inner self*], która mogłaby inteligentnie manipulować reprezentacjami. Postawiło go to w sytuacji, gdzie idee i wrażenia musiały zacząć myśleć same za siebie. Rezultatem tego była teoria jaźni jako wiązki wrażeń i idei. Wchodziły one w interakcje, łączyły się samoczynnie, bez udziału nadrzędnego w stosunku do nich, nadzorującego ich działanie podmiotu. Coś

[85] Tamże, s. 113.

takiego nie mogło funkcjonować, wrażenia i idee nie posiadały do tego celu dostatecznej ilości informacji.

Pojawia się więc problem regresu do nieskończoności, gdyż trzeba dodać choć jednego rozumiejącego interpretatora reprezentacji, a on pociąga za sobą kolejnych. Dennett stwierdza, że dopiero sztuczna inteligencja umożliwiła uniknięcie tego problemu. Wprowadziła ona pojęcie struktur danych [*data-structures*], które zdaniem Dennetta są funkcjonalnymi przykładami reprezentacji. W wymaganym tu sensie można o nich powiedzieć, że rozumieją same siebie.

Dennett wyjaśnia to zaczynając od poziomu osoby lub organizmu poznającego, który nazywa systemem intencjonalnym. Następnie rozbija ten system na organizacje podsystemów, z których również może być rozpatrywany jako system intencjonalny, ze swoimi przekonaniami i pragnieniami, a więc jako homunculus. Mówienie o homunculusach jest wszechobecne w sztucznej inteligencji i jest to tylko metaforyczny sposób wyrażania się. W systemach intencjonalnych istotna jest organizacja i współdziałanie wyspecjalizowanych homunculusów. Funkcja każdego z nich jest dalej dzielona na podfunkcje niższych homunculusów, aż do homunculusa, który zdolny jest „powiedzieć" tylko „tak" lub „nie", i którego zastąpić może maszyna.

Homunculusy działają, wysyłają wiadomości, a więc muszą mieć reprezentacje. Dla stosunkowo słabo zinterpretowanej reprezentacji, np. wzór na siatkówce, potrzebny jest bardziej złożony interpretator i użytkownik. Jeżeli interpretacja zawiera więcej informacji, np. co do związanego z nią postępowania, jej użytkownik może być mniej złożony. Nie występują reprezentacje rozumiejące się całkowicie samodzielnie, poza sytuacją, kiedy z globalnego punktu widzenia obserwowane są wszystkie reprezentacje systemu intencjonalnego. Pojedyncze homunculusy nie mogą zastąpić reprezentacji, czy zdolności językowych człowieka jako całości.

Jeżeli przyjmie się punkt widzenia, że struktury danych nie są reprezentacjami, to zdaniem Dennetta sukcesy sztucznej inteligencji świadczą o tym, że psychologia nie potrzebuje wcale reprezentacji i może posługiwać się takimi pseudoreprezentacjami, jakimi są struktury danych. Dennett twierdzi, że są one jedynym znanym rozwiązaniem problemu Hume'a.

Intencjonalność, w ujęciu Dennetta, może być realizowana przy udziale wielkiej, ale skończonej liczby homunculusów, które można też nazwać modułami operacyjnymi, np. moduł pamięci, rozpoznawania figur, dokonywania operacji arytmetycznych itp. Żaden z tych modułów nie posiada wewnętrznie tej treści intencjonalnej, jaką posiada człowiek jako całościowo potraktowany system intencjonalny. Homunculusy mają wewnętrznie struktury danych nazywane reprezentacjami lub pseudoreprezentacjami[86]. Człowiek jako całość nie musi tych cząstkowych intencjonalności posiadać wewnętrznie, w sensie uświadamiania ich sobie. Posiada je natomiast w sensie fizykalnego posiadania mózgu wraz z jego strukturami neuronalnymi. Zaczynając od tego poziomu wyodrębniać można coraz bardziej złożone homunculusy, posiadające wewnętrznie pewne treści intencjonalne lub struktury danych, aż do poziomu człowieka jako całości.

[86] Inaczej uważa J. Searle. Stwierdza on, że treść intencjonalna jest wewnętrzna w stosunku do stanów intencjonalnych. Dzięki intencjonalności umysł ma zdolność reprezentowania. Dla istnienia reprezentacji potrzebny jest jakiś podmiot, który używa pewnej wyodrębnionej całości – obrazu, zdania lub innego przedmiotu – jako reprezentacji. Teoria taka wymaga homunculusów z ich własną intencjonalnością, co wywołuje problem regresu do nieskończoności.

4.4. Maszyny myślące i odczuwające

Mechanistyczną teorię umysłu można by odrzucić, gdyby udało się pokazać, że umysł ludzki potrafi dokonywać rzeczy, których nie potrafią maszyny. Istnieje rodzina argumentów odwołujących się do twierdzenia Gödla[87], które przytacza Dennett. Jego zdaniem argumenty te nie są trafne, ponieważ pomijają oczywistą prawdę mówiącą, że logika nie skupia się bezpośrednio na świecie, lecz na opisach lub interpretacjach rzeczy. Argument antymechanistyczny sprowadza się do tego, że każda maszyna komputacyjna może być reprezentowana przez pewną maszynę Turinga, ale nie może być przez nią reprezentowany człowiek. Błąd tego argumentu tkwi zdaniem Dennetta w założeniu, że ograniczenie aktywności i zdolności człowieka może przebiegać w pewien neutralny sposób, nie odwołujący się do stosowalności twierdzenia Gödla. Twierdzenie to mówi, że w każdym niesprzecznym systemie aksjomatycznym, wystarczająco bogatym, by wygenerować w nim arytmetykę liczb naturalnych są twierdzenia, których nie można udowodnić. Taki abstrakcyjny system może zostać potraktowany jako instrukcje dla pewnej maszyny. Również maszynę Turinga uznać można za skończony zbiór instrukcji, lecz instrukcje te muszą mieć swoja realizację w mechanicznym lub elektronicznym urządzeniu. Fizyczne podłoże takich instrukcji musi uwzględniać wszystkie różnice stanów i użytych symboli. Maszyna Turinga, dla której istotne są tylko zależności syntaktyczne, może być wykorzystywana do wielu celów. Zależą one od tego, jakie znaczenia i jaką semantykę nadadzą symbolom tej maszyny jej projektanci i użytkownicy.

Dennett rozważa możliwość istnienia żywych maszyn Turinga, przy czym pojawia się problem nadania znaczenia stosowanej przez nie symbolice. Konieczne jest w tym celu potraktowanie ich zachowań w sposób mechaniczny, określenie informacji na wejściu i wyjściu oraz wskazanie błędów w działaniu. Potraktowanie w ten sposób człowieka wymaga przypisania jego zachowaniom określonych stanów maszyny Turinga. Nie będzie tu jednak projektanta, którego można by spytać o znaczenie stosowanych przez maszynę symboli, a jedynym jej użytkownikiem będzie zinterpretowany w taki właśnie sposób człowiek. Twierdzenie Gödla dotyczy go tylko o tyle, o ile jest on realizacją pewnej maszyny Turinga, które kolejne stany potraktowane zostały jako kroki dowodu tego twierdzenia. Błędem jest, zdaniem Dennetta, zakładanie jednej określonej maszyny Turinga dla danego podmiotu, by w filozofii umysłu stosowalne było twierdzenie Gödla. Postępowanie takie nie wnosi wiele nowego do badań nad zasadami funkcjonowania umysłu.

Zdolność mowy porównuje Dennett do funkcji komunikowania przez maszynę jej stanów wewnętrznych i rozważa sposoby kontrolowania mowy przez mechanizmy neuronalne. Wypowiedź informująca o stanie wewnętrznym nie jest bezpośrednią funkcją danego stanu logicznego, lecz wynikiem pracy systemów, do których mamy tylko ograniczony dostęp. Pewne szczegóły dotyczące systemów kontroli mowy można jednak, zdaniem Dennetta, wywnioskować ze struktury języka. Zgadza się on z Chomskym, co do tego, że struktury kontrolujące zachowania językowe muszą być raczej hierarchiczne niż szeregowe[88]. Kontroli podlega całe zdanie lub cała wypowiedź, następnie jej poszczególne części, nie tylko same części z osobna.

Nie znaleziono dotychczas żadnych anatomicznych odpowiedników tej hierarchii, ale zakładanie jej jest pomocne w umiejscawianiu błędów działania systemów odpowiedzialnych

[87] Dennett powołuje się na: J.R. Lucas, *Minds, Machines and Gödel*, w: "Philosophy", XXXVI/1961, s. 112 – 127.

[88] Por. D.C. Dennett, *Content and Consciousness*, s. 104.

za komunikowanie stanów wewnętrznych. Nie można jeszcze dokładnie określić miejsca ich występowania w mózgu, ale przyjmowanie takiej hierarchii jest, zdaniem Dennetta, właściwym sposobem ich poszukiwania. Dennett nie uznaje jednak identyczności myśli z procesami mózgowymi, ponieważ nie dysponuje mechanizmem uświadamiania tych procesów. Popełniane przez nie błędy podzielić można na dwie grupy. Zdarzają się one w poprzedzających centrum mowy systemach aferentnej analizy lub już w samym centrum. Te ostatnie powodują przejęzyczenia. Błędy dotyczą tego, co ma być wyrażone, w pierwszym wypadku, lub tego jak ma to być wyrażone, w wypadku drugim. Nie ma miejsca na błędne zinterpretowanie tego, co pojawia się na wejściu centrum mowy. Bez względu na to, czy docierająca tam informacja zawiera błąd, czy też nie, komunikat dotyczy stanu na wejściu, a nie bezpośrednio świata zewnętrznego. Wypowiedziane zdanie nie jest opisem zdarzenia cerebralnego, lecz wyrażeniem jego treści, która w szczególnym przypadku może tego zdarzenia dotyczyć.

Powracając na poziom treści Dennett nie dodaje nic nowego do wyjaśnień problemu połączenia myśli ze zdarzeniami neuronalnymi. Wydaje się, ze chce on uniknąć tego problemu i zbadać tylko możliwość projektowania i konstruowania modeli, które doskonale będą naśladowały zachowanie się ludzkich umysłów. Nigdy nie uda się jednak stwierdzić, czy model taki jest rzeczywiście świadomy, czy też tylko bycie świadomym, w sensie, w jakim świadomi są ludzie, udaje. Model taki zawsze będzie umysłem sztucznym i to może być przyczyna pominięcia pewnej istotnej cechy umysłu, który powstał naturalnie na drodze ewolucji. Dennett nie jest jednak zwolennikiem witalizmu i uważa, że problem tkwi w rozpoznaniu i odwzorowaniu wszystkich spełnianych przez umysł funkcji, na dowolnym, przydatnym do tego celu podłożu fizycznym.

Obecnie komputery zdolne są wykonywać wiele skomplikowanych czynności, takich jak gra w szachy, czy prowadzenie dialogu w dowolnym języku naturalnym. Dennett przytacza rozróżnienie Keith Gunderson[89] na sukcesy sztucznej inteligencji związane z intelektualnymi [*sapient*] cechami mentalnymi i porażki związane z jej cechami zmysłowymi [*sentient*]. Cechy mentalne podzielone zostały na te, które poddają się odwzorowaniu w programie komputerowym i te, które stawiają temu opór. Wskazuje to na odmienność problemu zaprojektowania maszyny myślącej od problemu zaprojektowania maszyny odczuwającej. Wnioskować można, że zrealizowanie sztucznego czucia zależy od znalezienia właściwego *hardware*, nie od napisania właściwego programu.

Symulacje komputerowe pozwalają przewidywać skutki wydarzeń, nie można jednak dzięki nim skutków tych odczuć. Komputerowa symulacja bólu jest, zdaniem Dennetta, tylko symulacja zewnętrznego, związanego z bólem zachowania. Można symulować również zachowanie się włókien nerwowych uczestniczących w przewodzeniu związanego z bólem sygnału. W ten sposób realizuje się sztucznie związki przyczynowo – skutkowe biorące udział w powstawaniu bólu, ale nie sam ból. Zwolennik teorii identyczności mógłby powiedzieć, ze bólem jest pobudzenie pewnego włókna nerwowego. Sceptyk nalegać będzie, że nie wyjaśnia to wcale pojawiania się jakości bólu. Wyjaśnienie takie można podać poprzez zsyntetyzowanie prawdziwego bólu, a nie poprzez jego symulację w programie komputerowym.

Ból może być rozważany jako pewien proces, zdarzenie, stan, rodzaj biologicznej lub psychologicznej substancji. Zależy to od zastosowanego sposobu użycia rzeczownika „ból".

[89] Por. D.C. Dennett, *Brainstorms*, s. 190.

W wyrażeniu „Czuję ból” jest jeden, policzalny ból, a w stwierdzeniu „Spowoduje to więcej bólu” ból traktowany jest jako substancja o nieokreślonej ilości. Dla potwierdzenia ilościowej i policzalnej natury bólu Dennett podaje przykład stosowania bólu, jako antidotum na przedawkowanie morfiny[90]. Podmiot cierpiący większy ból może przyjąć i zneutralizować większą jej ilość, niż ktoś doświadczający tylko małego bólu. Zastanawiając się nad tym, czy sztuczny ból byłby taki sam jak ból prawdziwy Dennett prowadzi analizy językowe. Zwraca uwagę na równoległe funkcjonowanie terminów: sztuczny [*artificial*] i syntetyczny [*synthetic*]. Użycie ich nie zawsze jest zamienne. Sztucznie uzyskane kolory są takie same jak kolory naturalnie występujące w przyrodzie, ale sztuczne kwiaty już nie. Nie przyjęło się też nazywanie ich kwiatami syntetycznymi. „Sztuczne” lepiej oddaje ich niedoskonałość. Sztuczny, czy też syntetyczny ból i sztuczna inteligencja mogą, zdaniem Dennetta, zastąpić ich naturalne odpowiedniki, jeżeli tylko zaprojektowane zostaną jako spełniające dokładnie takie same funkcje. Jedynym wypadkiem, w którym takie zastąpienie okaże się niemożliwe będzie wykazanie czysto biologicznej natury fenomenów bólu i inteligencji.

Naturalny ból związany jest z przetrwaniem i podatnością organizmu na uszkodzenia. Pojmowanie bólu może być również związane z intuicjami etycznymi dotyczącymi zła i cierpienia. Może to powodować trudności w wytworzeniu sztucznego, czy syntetycznego bólu, gdyż zagadnienia te nie poddają się całkowitej kontroli komputerowej. Dennett zauważa, że tkanki organiczne mogą okazać się jedynym podłożem, na którym możliwe jest zrealizowanie wymaganych dla bólu struktur funkcjonalnych. Jeżeli do zrealizowania tych struktur posłuży pewien materiał nieorganiczny, to nie powstanie w tym wypadku prawdziwy ból, gdyż istotna może okazać się biochemia stanu bycia w bólu. Wymagania takie stawiane są przez witalistów, a zdaniem Dennetta możliwość lub niemożliwość znalezienia odpowiedniego dla spełniania struktur funkcjonalnych bólu, sztucznego lub syntetycznego podłoża nie została jeszcze przesądzona.

Dennett przedstawia szkic, który nie jest poważnie potraktowanym programem komputerowym, ani poważną koncepcją psychologiczną lub neurofizjologiczną, lecz ma ułatwić filozoficzne badanie problemu bólu. Analizę anatomii bólu zaczyna od peryferyjnych powierzchni skóry, skąd receptory bólowe, za pośrednictwem dwóch rodzajów włókien nerwowych, przesyłają sygnały do mózgu. Pokryte mieliną włókna A przekazują sygnał z szybkością 100 metrów na sekundę, a cienkie włókna C, pozbawione powłoki mielinowej, zdolne są do przekazywania sygnału tylko z szybkością mniejszą niż dwa metry na sekundę. Można przypuszczać, że włókna C uczestniczą w przekazie „tępego”, „głębokiego” bólu, a włókna A w przekazie bólu „ostrego” i „palącego”. Dennett przytacza sugestię Melzacka i Walla[91], którzy przypisali włóknom A funkcję zamykania przejścia lub zmniejszania wartości wyjściowej na drodze przekazu sygnału bólowego poprzez włókno C. W ten sposób ograniczony sygnał poprzez zaproponowane przez Melzacka i Walla wejście dociera do wzgórza [*thalamus*] i rozdzielony zostaje następnie na dwa kanały. Jeden z nich biegnie do filogenetycznie starszej części mózgu, a drugi do kory mózgowej, która jest stosunkowo słabo rozwinięta u niższych gatunków zwierząt. Kora mózgowa wykorzystywana jest w procesach szczegółowych percepcji, a także umiejscawiania oraz charakteryzowania różnych odmian bólu i innych bodźców.

Dalsza analiza, którą przeprowadza Dennett, nie uwzględnia struktur anatomicznych, a tylko abstrakcyjnie wydzielone podsystemy. Zaproponowane zostaje istnienie pewnych filtrów kontrolujących natężenie sygnału. Działanie ich ma polegać na przykład, na obniżaniu

[90] Tamże, s. 196.
[91] Tamże, s. 200.

poziomu bólu poprzez skupianie uwagi na czymś z bólem niezwiązanym. Natomiast reakcje odruchowe organizmu świadczą o istnieniu bezpośrednich połączeń receptorów bólowych na wejściu systemu odpowiedzialnego za reakcje na ból, z podsystemami odpowiedzialnymi za skurcze mięśni, z pominięciem wyższych systemów kontroli w mózgu. Istotną rolę spełnia, jak pisze Dennett, umiejscowiony w pniu mózgu, w tworze siatkowatym [*reticular formation*], system aktywacji tworu siatkowatego [*RAS – reticular activating system*]. Odpowiedzialny jest on za czuwanie, sen i skupianie uwagi. Dodane do tego moduły: pamięci, przekonań, pragnień, analizy percepcyjnej i inne mają zapewnić niezawodną pracę systemu doświadczania i odczuwania bólu. Wydaje się jednak, że sceptyk, z którym Dennett dla zobrazowania swych poglądów prowadzi dialog, mógłby wciąż mieć wątpliwości, czy tak przedstawiony system funkcjonalny będzie wystarczający dla odczuwania, a nie tylko pojmowania jakości bólu. Skoro jednak odrzucone zostaje rozwiązanie leżące poza zasięgiem możliwości poznawczych człowieka, problem musi zostać sprowadzony do kwestii znalezienia odpowiedniego tworzywa [*hardware*] i programu [*software*]. Jak dotychczas zadania te zostały wykonane tylko na drodze naturalnej ewolucji, lecz zdaniem Dennetta badania związane ze sztuczną inteligencją, psychologia poznawczą i filozofia umysłu są na drodze obiecującej możliwość powtórzenia tych osiągnięć w sposób sztuczny.

Koncentrowanie się na własnym bólu może zmniejszyć jego natężenie i Dennett proponuje hipotezę, że zwiększa ono natężenie sygnału włókna A, które w teorii przekazu Melzacka i Walla ma działanie hamujące. W ten sam sposób działać ma dotykanie bolących miejsc. Ból odczuwany jest w określonych miejscach, choć zdarzają się przypadki odczuwania bólu w amputowanej kończynie. Nie występują one jednak, jeżeli amputacja dokonana została zaraz po urodzeniu, co świadczy o tym, że umiejscawianie bólu jest wyuczone. Natomiast podanie morfiny w momencie, kiedy ból jest już odczuwany powoduje, ze podmiot dalej odczuwa ten ból i identyfikuje go właśnie jako ból, lecz nie powoduje to cierpienia. Proponuje, aby wyjaśnić to w ten sposób, że kiedy sygnał mówiący o bólu dostaje się do modułu analizy percepcyjnej zostaje on rozpoznany, lecz morfina odcina połączenia nerwowe prowadzące do filogenetycznie starszej części mózgu i odczucie cierpienia zostaje zablokowane. Na podstawie tego, że kurara posiada właściwości czasowo paraliżujące, lecz nie odbierające świadomości i zdolności odczuwania oraz tego, że niektóre środki stosowane w anestezji powodują amnezję, rozważa hipotezę, w której nie ma utraty świadomości, a tylko paraliż i następujące po nim usunięcie z pamięci wszelkich wspomnień dotyczących danego okresu czasu[92]. Paradygmat świadomości jako warunku koniecznego dla odczuwania bólu nie ma więc statutu hipotezy sprawdzonej empirycznie. Niemożliwe jest odróżnienie braku świadomości od paraliżu połączonego z amnezją. Nie można wykazać, że ból zanika wraz z zanikiem świadomości, gdyż nie ma sposobu sprawdzenia, czy podmiot przestaje być świadomy, czy tylko traci zdolność uzewnętrzniania swego stanu, a następnie doznaje amnezji.

Możliwość doznania pewnego znieczulenia na ból dzięki skupianiu uwagi na pewnych treściach świadczy, zdaniem Dennetta, o udziale kory mózgowej w odczuwaniu bólu. Kora mózgowa odpowiedzialna jest za wyższe funkcje psychiczne, w niej zachodzą zdarzenia neuronalne odpowiednie dla danych treści, które mogą odwracać uwagę podmiotu od bólu. Niższe gatunki zwierząt, pozbawione kory mózgowej, odczuwają więc ból inaczej niż ludzie. Nie używają one języka werbalnego, a odczucie, którego doznają pozbawione jest związków z treściami kultury, w jakie wyposaża je człowiek. Pozostaje jednak filogenetycznie starsza część mózgu i samo czyste odczucie.

[92] Tamże, s. 210.

Dennett dokonując tego rodzaju analiz neurofizjologicznych na poziomie subosobowym zauważa, że pomijają one problem odczuwania jakości bólu. Subosobowy poziom wyjaśniania nie wnosi, jego zdaniem, rozwiązania filozoficznego problemu bólu. Wprowadzenie do proponowanego modelu homunculusa bólu spowoduje tylko to, że ból odczuwany będzie przez tego homunculusa, a nie przez podmiot jako całość. To nie skóra, czy kończyna odczuwa ból, ale cały podmiot. Ból musi więc być funkcją lub kombinacją wszystkich elementów składowych modelu. Nie ma jednak żadnej, wystarczająco dobrej teorii bólu. Nie można zaprojektować i skonstruować odczuwającego ból robota, gdyż nie wiadomo jaką teorie bólu miałby projekt tego robota uwzględniać. Jeżeli tylko podana zostanie wymagana fizjologiczna teoria bólu na poziomie subosobowym, możliwe będzie, zdaniem Dennetta, skonstruowanie robota odczuwającego ból. Problemem jest powiązanie poziomu subosobowego, na którym ból nie występuje i zachodzą tylko oddziaływania przyczynowo – skutkowe poszczególnych modułów operacyjnych, z poziomem osobowym, na którym pojawia się jakość bólu.

Roboty mogą stać się pewnego dnia tak samo świadomymi i odczuwającymi ból jak ludzie. Badania organizmów żywych pokazują, że są one pewnego rodzaju mechanizmami i dlatego rozważana jest możliwość utworzenia życia jako artefaktu. Pomysł ten ma wielu przeciwników. Sceptycyzm ich nie opiera się jednak na założeniu, że pojęcie świadomych maszyn jest sprzeczne z prawami natury, lecz na tym, że wyewoluowały one i są zbudowane na podstawie niezwykle złożonych projektów, których umysł ludzki nie potrafi odgadnąć. Jednak zdaniem Dennetta, zaprojektowanie maszyny obdarzonej świadomością leży w zasięgu ludzkich możliwości, a problemem jest tylko realizacja takiego projektu. Świadome roboty będą prawdopodobnie zawsze zbyt kosztowne. Dennett opowiada się jednak za możliwością wytworzenia ich jako artefaktów. Przypomina on charakterystyczny dla dualizmu problem wzajemnych oddziaływań: fizykalnego mózgu i niefizykalnego umysłu. Odrzucając dualizm można jednak utrzymywać, że świadomość występuje tylko w przypadku organicznych mózgów, a roboty są nieorganiczne. Tego rodzaju witalizm zostaje przez Dennetta odrzucony, gdyż biochemia wykazuje, że organiczne składniki istot żywych są redukowalne do mechanizmów i mogą być realizowane na innych niż organiczne, nośnikach fizykalnych. Nie mogą jednak zostać w ten sposób zrealizowane wszystkie funkcje ludzkiego, opartego o organiczny mózg, umysłu. Dennett zauważa, że szybkość i zwartość procesów mózgowych nie dają się zrealizować na innego rodzaju nośniku. Konieczne jest zastosowanie sztucznie wytworzonych, na bazie polimerów, tkanek organicznych. Zdaniem Dennetta żadne poważne filozoficzne i naukowe tezy nie mogą wiązać się z tezą, że może zostać skonstruowany nieproteinowy [*protein-free*], świadomy robot.

Możliwe, że roboty jako artefakty nie będą mogły posiadać świadomości, gdyż powstaje ona tylko na drodze naturalnej. W przekonaniu Dennetta jest to jedynie problem idealnej imitacji, która może zostać zrealizowana przy zapewnieniu wystarczających środków. Roboty mogą jednak pozostać na zawsze zbyt proste, by być świadomymi. Człowiek składa się z miliardów komórek, a pojedyncza komórka jest mechanizmem, którego skonstruowanie wciąż stanowi problem dla pracujących nad tym specjalistów. Możliwe jest jednak zbudowanie sztucznego serca, oka, czy ucha. Nie są one tak skomplikowane jak organy naturalne i nie działają tak jak one dokładnie. Zbudowanie ich jest jednak możliwe. W zasadzie możliwe jest również zbudowanie sztucznego mózgu lub różnych jego części. Wymaga to jednak obszernych badań i wielkiego nakładu pracy.

W laboratorium sztucznej inteligencji w MIT, pod przewodnictwem profesorów: Rodneya Brooksa i Lynn Andrea Stein podjęto próbę stworzenia najbardziej jak dotychczas

członekopodobnego robota o imieniu Cog. Również profesor Dennett uczestniczy w tym programie badawczym. Pisze on, że Cog jest wielkości dorosłego człowieka, ma dwa ramiona o długości ramion ludzkich i parę oczu mogących dokonywać trzech fiksacji na sekundę, podczas gdy oczy ludzkie dokonują ich cztery lub pięć. Cog posiada w pamięci program zawierający informacje dotyczące jego symulowanego dzieciństwa. Wyposażony jest w wiele systemów elektronicznych spełniających funkcje analogiczne do funkcji spełnianych przez różne części układu nerwowego człowieka, jak na przykład funkcja systemu rozpoznawania twarzy. System nerwowy Coga posiada architekturę równoległą, dzięki której jest on zdolny do równoczesnego przeprowadzania wielu operacji. Powstawać będą nowe generacje ulepszonych modeli tego robota, co ma naśladować proces ewolucji. Jednym z planowanych osiągnięć jest nauczenie Coga podstaw języka naturalnego, jednak nie poprzez wbudowanie mu specjalnego wewnętrznego organu językowego LAD[93], a poprzez naśladowanie utworzenia takiego organu na drodze trwającego tysiące generacji procesu ewolucyjnego. Cog posiada cztery mikrofony zastępujące uszy i ma mieć wbudowany syntezator mowy, a także odpowiednie oprogramowanie służące do analizy sygnałów dźwiękowych, wystarczające dla rozróżnienia dźwięków mowy ludzkiej. Ma to posłużyć sprawdzeniu, czy Cog będzie pierwszym robotem zdolnym samodzielnie uczyć się wysoce skomplikowanych struktur. Dla powodzenia tego projektu potrzebne jest również wyposażenie Coga w ludzkie dążenia, obawy i upodobania. Takie antropomorfizmy można, jak pisze Dennett, zastąpić innymi terminami. Mogą to być rejestry celów [*goal-registrations*] i funkcje preferencji [*preference-functions*], które są z tymi antropomorfizmami, z grubsza rzecz biorąc, izomorficzne.

Początkowe wyposażenie Coga, z którego rozwinięty ma zostać jego sztuczny system nerwowy, złożone zostało z 64 komputerów Mac II. Pracują one w architekturze równoległej i każdy z nich wyposażony został w multiprocesor oraz 1 megabajt RAM-u. Dzięki ich oprogramowaniu można uruchomić jakąkolwiek z wielu maszyn wirtualnych związanych z wyposażeniem genetycznym człowieka lub z jego symulacją. Maszyny wirtualne, które można uruchamiać w sztucznym mózgu Coga są jednak, jak pisze Dennett, niewielkie w porównaniu z możliwościami ludzkiego mózgu. Pomimo tego takie uproszczone projekty dają pewien wgląd w mechanizmy ewolucji. Wydaje się jednak, że proponowana przez Dennetta maszyna wirtualna nie wnosi wiele nowego do rozwiązania problemu świadomości. Można prześledzić przebieg pewnego wirtualnego procesu, ale w obrębie takiej maszyny nie ma możliwości sprawdzenia, czy ona sama jest świadoma. Stawianie takiego zarzutu maszynie może być jednak porównane do stawiania takiego samego zarzutu innym ludziom. Nie ma sposobu na sprawdzenie, czy inni są świadomi, czy też są tylko pozbawionymi świadomości automatami. Rozwiązaniem tego problemu może być tylko wynalezienie maszyny do „czytania" cudzych umysłów, w sensie wczytania się w specyficzny stan cudzej świadomości, a więc bycia kimś innym. Maszyna taka mogłaby zostać użyta również do „odczytania" umysłu maszyny. Ciekawe tylko, czy człowiek lub świadoma maszyna zgodziliby się na tak dalekie naruszenie własnej prywatności? Poza tym gdyby maszynę do „czytania" umysłów uczynić świadomą, mogłaby ona z pełną świadomością swego postępowania nadużyć możliwości wynikających z poznania wszelkich tajemnic, ukrytych w umysłach żywych i mogących podlegać „odczytaniu", przedstawicieli rodzaju ludzkiego.

[93] Z języka angielskiego: *Language Acquisition Device*.

5. Zakończenie

Praca ta nie wyczerpała z pewnością całości poglądów Dennetta, lecz przedstawiła jego główne obszary zainteresowań. Może ona być dobrym wprowadzeniem do dokładniejszego poznania proponowanej przez niego koncepcji oraz przeglądem wielu podejmowanych współcześnie zagadnień sztucznej inteligencji.

W przedstawionej tu naturalistycznej koncepcji umysłu, połączone zostało filozoficzne pojęcie intencjonalności z naturą i organizmami żywymi (ewolucja), z jednej strony, oraz z kulturą i osiągnięciami najnowszych technologii (sztuczna inteligencja), z drugiej. Intencjonalność pojęta została tutaj bardzo szeroko. Najbliższe intencjonalności pierwotnej i wewnętrznej jest nakierowanie prostych organizmów żywych na przetrwanie. Wszelkie inne intencjonalności to tylko intencjonalności pochodne. Niesłuszne jest więc przekonanie o istnieniu w umyśle człowieka intencjonalności pierwotnej i wewnętrznej.

Intencjonalność w tym ujęciu nie jest wynikiem działania inteligencji, ona wraz z inteligencją powstaje. Proces, w którym kształtuje się intencjonalność wraz z inteligencją to ewolucja gatunków poprzez dobór naturalny. Jego początkiem jest powstanie pierwszych organizmów żywych, które wyodrębniły swe wnętrza i ustanowiły granicę pomiędzy nimi, a światem na zewnątrz. Dennett nie jest zainteresowany warunkami początkowymi tego procesu, ani problemem samej możliwości pojawienia się intencjonalności i inteligencji we wszechświecie. Nie jest to potrzebne dla szkicowania projektów umysłów sztucznych. Odwołują się one jednak do poziomu fizykalnego (nastawienie fizykalne), którego Dennett nie wyposaża w intencjonalność pierwotną i wewnętrzną. Korzysta z dorobku nauk fizykalnych jak z gotowych podzespołów i czasem bardziej jest konstruktorem niż filozofem.

Można spróbować przenieść i tak już szeroko potraktowane przez Dennetta pojęcie intencjonalności, zarówno do skali mikro jak i makrokosmosu. Wtedy intencją sił wewnątrzatomowych będzie utworzenie układu okresowego pierwiastków, a intencją grawitacji dystrybucja materii we wszechświecie, w formie galaktyk i układów planetarnych. Jedno i drugie potrzebne jest dla powstania pierwszych organizmów żywych. W takim ujęciu intencjonalność nie pojawia się wraz z pierwszymi formami życia, a jest obecna we wszechświecie już od jego początku. Stworzenie warunków odpowiednich dla pojawienia się życia należy wtedy uznać za intencję wszechświata. Zastanawiające jest, czy zaistnienie takich warunków w pewnym miejscu we wszechświecie, może być tylko przypadkowe?

Dennett nie porusza problemów filozoficznych, które nie są związane z zagadnieniem sztucznej inteligencji. Jego zainteresowania skupiają się na poznaniu wszystkich funkcji składających się na świadomy ludzki umysł. Poznanie to ma służyć odtworzeniu projektu, którego autorem jest proces ewolucji poprzez dobór naturalny. Za jedno z największych osiągnięć teorii ewolucji, uważane jest usunięcie z niej inteligencji. Coraz lepiej dostosowane do środowiska organizmy żywe, u Dennetta systemy intencjonalne, powstają dzięki czynionym na ślepo, a więc nieinteligentnym próbom. W wyniku tego procesu powstają w końcu systemy intencjonalne na tyle inteligentne, by chciały zrozumieć zasadę swego powstania i wiele innych rzeczy. Zgodnie z teorią ewolucji ich inteligencja musiałaby powstać samoczynnie, na bazie pierwszych, nie będących skutkiem działania inteligencji posunięć, wzmacnianych i powtarzanych, jeżeli okazały się korzystne. Ciekawe, czy zaistnienie takiej sytuacji we wszechświecie może być tylko dziełem przypadku?

Poza pytaniami tego rodzaju interesujące są również przewidywania dotyczące skutków, z którymi trzeba będzie się liczyć, jeżeli koncepcja Dennetta i jej podobne okażą się słuszne. Zaprojektowanie i uruchomienie sztucznego umysłu, a więc i sztucznego człowieka będzie wtedy, tylko kwestią czasu. Można zastanowić się nad związanymi z tym problemami etycznymi. Jest raczej mało prawdopodobne, aby już pierwszy sztuczny człowiek był rzeczywiście dobrą kopią człowieka naturalnego. Czy należy więc produkować ułomnych ludzi? Innym problemem jest określenie statusu społecznego takich tworów. Może to doprowadzić do sytuacji, w której nasi potomkowie będą mieli getta cyborgów i społeczności naturalnych ludzi. Kolejny z problemów to systemy kontroli, bez których sztuczni ludzie mogą okazać się niebezpieczni dla swych twórców.

Możliwości, jakie otworzyłby przed gatunkiem ludzkim sukces sztucznej inteligencji są tak wielkie, że trudno jest nawet wszystkie je przewidzieć. Można jednak zastanowić się, jaki sens ma wytwarzanie sztucznych ludzi, jeżeli tak wielu rodzi się w naturalny sposób. Zarówno naturalnym jak i sztucznym ludziom, (jeżeli będą w pełni ludźmi, a takimi mają być), wypada przyznać takie same prawa. Bardziej użyteczne wydaje się w tej sytuacji, produkowanie prostszych automatów użytkowych. Samo poznanie struktury sieci neuronalnych, gwarantujących funkcjonowanie świadomego umysłu, może jednak zostać wykorzystane do uruchomienia umysłu, będącego dokładną repliką umysłu konkretnego żywego człowieka. Sztuczne neurony takiego umysłu będą nieśmiertelne, lecz brak organicznego ciała i związanych z nim typowo ludzkich odczuć, może być przyczyną różnic pomiędzy umysłem sztucznym a naturalnym.

Badania prowadzone w laboratoriach sztucznej inteligencji znajdują się wciąż w swym stadium początkowym, a związane z nimi problemy są motywem często podejmowanym przez literaturę science fiction. Ze względu na ich złożoną naturę oraz kosztowność eksperymentów, dopiero stosunkowo odległa przyszłość mogła będzie przynieść znaczący postęp.

Literatura:

[1] A.J. Ayer, *Filozofia w XX wieku*, PWN, Warszawa 1997.

[2] N. Block, *On a Confusion about a Function of Cousciousness*, w: “Behavioral and Brain Sciences”, 18(2)/1995.

[3] D.C. Dennett, *Artificial Life as Philosophy*, w: “Artificial Life”, t. 1, N 2/1994.

[4] D.C. Dennett, *Brainstorms*, The Harvest Press Limited, Sussex 1978.

[5] D.C. Dennett, *Content and Consciousness*, Routledge & Kegan Paul, London 1969.

[6] D.C. Dennett, *Consciousness Explained*, Little Brown & Company, Toronto 1991.

[7] D.C. Dennett, *Consciousness in Human and Robot Minds*, (referat wygłoszony na) IIAS Symposium on Cognition, Computation and Consciousness, Kyoto, September 1-3, 1994.

[8] D.C. Dennett, *The Intentional Stance*, The MIT Press, Massachusetts 1996.

[9] D.C. Dennett, *Making Tools for Thinking*, (referat wygłoszony na) Conference at Simons Fraser University, Feb. 7-8, 1997.

[10] D.C. Dennett, *Memes: Myths, Misunderstanding and Misgivings*, w: "Draft for Chapel Hill", October 1998.

[11] D.C. Dennett, *Natura umysłów*, Wydawnictwo CIS, Warszawa 1997.

[12] D.C. Dennett, M. Kinsbourne, *Time and the Observer*, w: "Behavioral and Brain Sciences", 15/1992.

[13] *Filozofia umysłu, Fragmenty filozofii analitycznej*, red. wyd. Bohdan Chwedeńczuk, Wydawnictwo Spacja, Warszawa 1995.

[14] J. Heil, *Philosophy of Mind*, Routledge, London 1998.

[15] D. Hume, *A Treatise of Human Nature*, Penguin Books, London 1984.

[16] Platon, *Teajtet*, PWN, Warszawa 1959.

[17] H. Putnam, *Wiele twarzy realizmu i inne eseje*, PWN, Warszawa, 1998.

[18] W.V.O. Quine, *Słowo i przedmiot*, PWN, Warszawa 1999.

[19] G. Ryle, *Czym jest umysł?*, PWN, Warszawa 1970.

[20] J.R. Searle, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press 1983.

[21] B.F. Skinner, *Science and Human Behavior*, The Free Press, New York 1965.